Discrimination of Neutral Postures in Computer Based Work

THESIS

Alanna R. Keith, Captain, USAF

AFIT-ENP-13-M-19

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENP-13-M-19

DISCRIMINATION OF NEUTRAL POSTURES IN COMPUTER BASED WORK

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Physics

Alanna R. Keith, B.S.

Captain, USAF

March 19, 2013

AFIT-ENP-13-M-19

DISCRIMINATION OF NEUTRAL POSTURES IN COMPUTER BASED WORK

Alanna R. Keith, B.S.
Captain, USAF

Approved:

Dr. Amy L. Magnus Chairman	Date
Dr. Gilbert Peterson Member	Date
Dr. Christoph Borel-Donohue Member	Date

AFIT-ENP-13-M-19

Abstract

Biometric authentication verifies users based on the way they physically interact with a system. In this thesis, we discover a neutral posture that typists consistently display during non-trivial computer work and explore its potential for distinguishing typists. We aim to demonstrate three objectives: first, compelling proof that a user can be actively verified over the course of a lengthy task via a neutral posture struck multiple times in the performance of that task; two, a sensing concept for capturing the neutral posture, and, third, an objective method for determine the level of work performed by each typist.

This thesis develops a method of hand tracking that uses a simple ellipse to model hand posture. Hand postures are tracked and characterized to distinguish a computer user's set position, the neutral posture where a typist pauses before typing. Initial results of a group of 10 users indicate that the neutral posture can be modeled based on only a couple of seconds of training data and that model can perform with approximately 92% accuracy. Our methods fuse overhead video with key logging data to achieve these results. Further, we estimated the complexity of the typists' work by aligning the verb phrases of the typed text with Bloom's Taxonomy—a taxonomy based on verb usage. Verb phrases indicate the level of competency that the user endeavored to demonstrate. This competency or expertise may further distinguish users and their performance in their most engaging work.

Acknowledgements

First of all, I would like to thank my research advisor, Dr Amy Magnus, for giving me this opportunity to expand my horizons into a field of research in which I had no experience going in and which, while not being the typical 'optics' type physics thesis, grabbed my interest from the moment we first listened to the thesis presentations at the beginning of this journey. I'd especially like to thank both her and LtCol Jeremy Holtgrave for all the help they've given me since the beginning.

Likewise, I'd like to thank my thesis committee members, Dr Gilbert Peterson and Dr Christoph Borel-Donohue, for giving me advice and lending me equipment. Their insight was extremely valuable.

I'd also like to thank Anum, for teaching me how to use all the fancy stuff in the VACE lab, Joe and Devin, for all the coding help you've given me, and Rebecca for letting me bounce ideas off you.

And finally, I'd like to thank Captain Stephanie Keith, also slaving away under the thumb of AFIT. Although I'm not entirely sure what I'm thanking her for, given that she was holed away in a sound and electronic proof room for the latter half of the thesis quarter conducting her own research and rarely answering her phone, I'm sure much thanks is due anyway.

Alanna R. Keith

Table of Contents

List of Figures

List of Tables

DISCRIMINATION OF NEUTRAL POSTURES IN COMPUTER BASED WORK

I. Introduction

1.1 Problem

This thesis investigates a common posture in computer work: the relaxed or neutral position of a hand at the computer before a user begins typing. Our goal is to understand this posture especially when a computer user is performing high level typing. Can a relationship be found between hand posture and higher order work? An authentication system must be able to verify computer users when they are performing at their highest level, as that is when they are producing critical work. These are times when the computer user absolutely does not want the computer system to question their access.

1.2 Research Objectives

A biometric authentication system should verify user identity based on several different modalities. We expect that each modality only functions well in a given range, and multiple modalities should be used to ensure thorough coverage and a more robust authentication system.

In the hand tracking modality, we can discover actions that are common among users and yet characteristic of individual users. In this research, we shall concentrate on the neutral posture that typists strike in order to characterize patterns in their behavior as they create a document. A pose struck between thought and action is called the 'set position'. The typist's set position is where their hands typically return

to after or just before a sequence of typing. The hands may also briefly move through this neutral position while the user is in the middle of typing.

The goal of this research is to characterize individuals by their set position. We will develop a simple mathematical model of the human hand that can be reliably fit to hands in video taken from a bird's eye view above the keyboard, and we will test whether the hand model can be formulated to distinguish between of multiple participants. Additionally, we wish to fuse together diverse data sources — video, text analysis, and keylogging data — to form a comprehensive model of a user's competency that may be used to determine their uniqueness.

1.3 Overview

The remainder of this document is organized as follows: first, a brief review of the previous research into computer authentication, biometric authentication systems, and tracking and modeling for both hands and fingers is presented, exploring data fusion techniques and characterizing phenomena. Next is a discussion on related work done at AFIT, followed by relevant theory, and then a discussion on the research approach. Finally, the results and conclusions are presented along with a proposal for future work.

II. Background

Computers are vulnerable systems. They store important information that can be critical to national security, and many missions within the Department of Defense and industry rely on computerized systems in order to function. A misplaced password, determined adversary, or careless employee can leave these systems exposed.

Common Access Cards (CAC) and passwords are the two computer authentication methods most widely used by the DOD. Every employee has a CAC and usually one or more passwords for computer systems that they work on. Generally, the more secure a system is, the longer and more complicated a password must be, and the more often the password must be changed over the course of a year.

Requiring long passwords leads to passwords that are either written down or forgotten after periods of nonuse. If proper password procedures are followed, different user names and passwords are used for every online system that a user accesses. As a consequence, users confuse or forget passwords, and increasingly succumb to the desire to write them down. Once written down, these password lists can be lost or stolen.

Common Access Cards also have risks. CACs can be left in machines by complacent users, or misplaced. Although passwords and CACs were intended as methods to increase authentication security, their complicated nature leaves vulnerabilities open to exploitation.

In addition to CACs and passwords, there are some less common authentication methods. These methods include face, fingerprint, and voice recognition and have found their way into industry and consumer devices. These methods are generally thought of as more secure than password protection since they cannot be written down like passwords. However, these biometric systems often require the computer user to hold still while submitting to the authentication procedure.

Although these less common methods are designed to supplement the authentication provided by a password, and in some cases are used in place of a password, they can still be circumvented. Some of these systems can be fooled by a simple photograph of a user, or a recorded voice.

The methods of observation for the aforementioned authentication devices are known as modalities. Attackers can more easily circumvent a system based on one modality than a system based on two or more. In a system with more than one modality, each modality must be dealt with correctly in order to access the system — a failure to produce the correct input for even one modality prevents system access.

A different approach to computer authentication might be similar to the Google search engine, which models a user's preferences and favored way of interacting with its Graphical User Interface, or GUI. As a user searches for things or interacts with any Google application, the search engine suggests offerings based on the apparent user's current and historical interactions. A similar authentication system would remember how a user interacts with its GUI and create a model for what that user is likely to do. When the user does something unexpected that doesn't fit the system's model, that event can trigger the system to examine whether he or she is the same person.

This type of system is a behavior based biometric authentication system. Behavior based authentication systems recognize a user based on the way the user physically interacts with the system without passwords, a CAC, or other disruptive authorization procedures. These systems continuously verify as the user works, rather than relying on a single authentication event.

Behavioral biometric authentication systems have the potential to be much more secure than the systems discussed above. When users walks away from a computer they have been using and then come back, the computer registers the inactivity and

the return to activity, and continues the verification process to check if the original user has returned. In contrast, when users step away from a typical computer system they were using after logging in with a CAC or a password, that system would remain vulnerable unless the user locked it or until it locks automatically.

Biometric authentication procedures have the potential to be less disruptive because the individual person is the pass key. No memorized passwords are required. The user simply starts interacting. Any person who obtains the correct credentials can access a typical password or CAC based computer system, but a user is much more difficult to accurately and continuously duplicate than a password or access card.

There has been a great deal of research into alternative methods of computer authentication, including biometrics. This chapter will go into a brief overview of recent research but is not meant to be an exhaustive literature review. In addition, this chapter will touch briefly on current work in hand tracking, data fusion, and behavior characterization projects.

2.1 Recent Research into Computer Authentication

According to Wiedenback et al, "Authentication is the process of determining whether a user should be allowed access to a particular system or resource" [1].

Traditional alphanumeric passwords are common authentication measures for computers, yet by their nature, they can create a security hole in computer authentication. Password protocol [1] [2] [3] generally states that such passwords should be 1) easy to remember, yet should also be random and hard to guess, 2) changed frequently, 3) different for each account, 4) never written down, 5) contain a mixture of letters of different case, numbers, and special characters, and 6) be at least 8 characters long. More secure systems may require passwords at least 15 characters long. This 'easy

yet complex' contradiction can lead to users re-using passwords, simplifying them, or keeping a physical list.

Clearly, an alternative is needed. Two areas of research providing authentication options are graphical passwords, which will be touched on briefly, and biometrics, which is closely related to this paper's research.

Graphical Passwords

Graphical passwords use an image to form the password [1] [2]. In the system PassPoints [1] [2], the user clicks on any location within an image, and a sequence of clicks forms the password. The system encrypts the password and calculates a region tolerance about the chosen click points that the user must click within when logging in. These types of passwords are not biometric, but they may offer better remembering potential than typical alphanumeric passwords, since the user does not need to memorize a complicated string of characters.

Another graphical password method [3] requires users to write or draw their own password as either characters or a simple image. Such graphical passwords may be difficult to implement a typical dictionary attack against, and again, may be more easily remembered.

Biometrics

Gestures may present a viable method for secure authentication on gesture-enabled devices. Memon, et al [4] [5] have created software where users can log into their iPads using hand gestures. One of these gestures involves physically turning the image of a combination lock on the iPad touch screen. Another software, iSignOn, available for the iPhone from Apple's App Store, requires a user to sign with his or her finger. These gestures work because each person's hand size, fingers, and how they place

them on the screen are all different, including the precise way they make the gesture. This combination of physical features and gestures works as an effective password that is very difficult to duplicate, even if another person knew the user's number combination [4] [5].

Biometric data based on features (face, fingerprints, etc.) must be stored securely. Biometric data cannot be easily replaced if compromised (as a user could replace a password or CAC), however, this type of data is noisy and difficult to use with traditional cryptographic techniques [6]. Memon, et al [6] investigate this problem and employ a 'secure sketch' and geometric transformation to encode the data.

2.2 Recent Research into Hand Models and Tracking

An online search reveals many methods for hand capture, the majority of which are done facing the camera and without object interaction. Many do not include finger tracking, which is still in its infancy. The following is a brief review of current hand modeling and tracking research.

Jmaa, et al [7] developed an approach for digit (finger) recognition from hand gestures. Hand detection and isolation is performed, followed by finger extraction by removing the palm. This approach is invariant to scale, rotation, and translation of the hand.

Manresa, et al [8] are developing a method for video game control using hand segmentation, tracking, and gesture recognition. Segmentation is performed with a learning algorithm using HSL (hue, saturation, and value). To add robustness to segmentation errors, a hand tracking algorithm is introduced that attempts to maintain and predict the hand state over time. Gesture recognition is performed via contour and ellipse approximation of the hand.

Hand tracking is also used to recognize sign language [9]. Tracking is accomplished

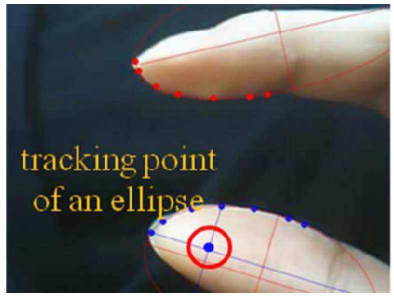


Figure 1. Credit to Kim, et al [14]. The form of the thumb and pointer finger are captured in the Active Shape Model (dots) and the ellipse being tracked.

using forward-backward prediction, and by incorporating statistical information. This tracking also functions during occluded cases, where the head and hands overlap. The occluded cases employ an ellipse to roughly denote the hand's position and contour.

Barhate, et al [10] have developed their Predictive EigenTracker to accurately track left and right hands during both occlusion and collision — instances where hands change their direction of motion during an occlusion. The EigenTracker can account for translation, scaling, and shear.

Rhee, et al [11] developed a method for constructing a person-specific three dimensional hand model from a single palm image of the hand without human guidance, based on feature extraction of creases on the palm and associated joint locations. A generic 3D hand model is then deformed using the features and contours of the hand image. The researchers noted that the three principal creases on the palm (distal palmar, proximal palmar, and thenar creases) are unique and may be suitable for biometric identification of a person [12] [13].

Fingertip modeling and tracking presents difficulties in that fingertips are small features and often occluded during gestures. Kim, et al [14] present a method for tracking using an Active Shape Model, which finds the shape of the fingertips. An ellipse is fitted to the model (Figure 1), and the fingertip is tracked via the ellipse (Figure 2).

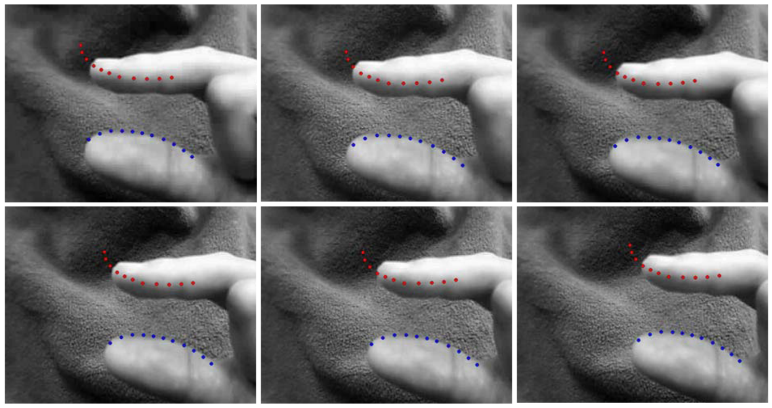


Figure 2. Credit to Kim, et al [14]. Image sequence showing tracking of fingertips as they move.

Candescent NUI [15], developed by Stefan Stegmueller for the Kinect, tracks both hands and fingertips using the Kinect's depth sensor. Originally planned for use in this thesis, the software functions best when the hands are not interacting with objects. Our early attempts used the Kinect's RGB sensor to perform background subtraction and create a hand contour to replace Candescent's own hand contour, which was formed using depth data. Despite these attempts to integrate hand tracking and finger identification against a keyboard into Candescent's code, the software was unable to reliably distinguish hands and fingers from the keyboard background. This difficulty pointed to both an inability of the Kinect depth sensor to adequately resolve touching objects through distance and to a lack of full understanding of the complex code.

Also problematic was the low temporal resolution of the Kinect. The advertised maximum resolution is 30 fps, but in practice, due to the process of saving each image frame in turn before the next frame can be read in and saved, the actual fps was lower depending on computer speed. A high temporal resolution is necessary when studying the short, quick movements of fingers on a keyboard.

2.3 Machine Learning

Allen, et al [16] developed PLOW, an advanced intelligent assistant that learns a task from a single learning session consisting of demonstration and speech. These assistants are systems that can interact with people and help them to perform everyday tasks. PLOW learns information management tasks that can be performed within a web browser, and users can interact with the system through either speech or text. The natural language understanding is accomplished with the TRIPS system [17].

Ferguson, et al [18] expand on the work by integrating natural language with different data sources that record physical human behavior. Kinect RGB and depth data, HD video, speech, and RFID data are integrated in order to allow a computer to learn what a person is doing during a task and to eventually duplicate that task or perform a slightly different but similar task. In comparison, we are working to integrate typing data, text, and HD video in order to recognize the differences between people based on their apparent competency.

Swift, et al [18] intend to use the Kinect depth data to help segment humans and objects. However, we've found that the hand was hard to discern when it is interacting with an object (keyboard), so we will continue to follow this work to see how they approach this problem.

2.4 Characterizing Behavior

The Defense Advanced Research Projects Agency (DARPA) has several initiatives involved in characterizing dismounted behavior. The VIRAT project, Video and Image Retrieval and Analysis Tool [19], seeks to recognize human actions in video and annotate the video appropriately, providing real-time actionable information as events unfold, and a way to search through archived video to retrieve content of interest [20].

While VIRAT attempts to characterize behavior in terms of temporal phenomena (starts and stops) and synthesis events (splits and joins) [19], it does not seek to establish or categorize based on apparent competency. Bloom's Taxonomy, discussed in Chapter 4, suggests that we can categorize behaviors by apparent competency, thereby establishing persons of interest. Bloom's taxonomy tracks competency, via categories from simple to complex, in human interactions with information, human to human interactions, and human interactions with physical interfaces. The Set posture is a common posture associated with human/interface interactions. We seek to tie this posture to the competency with which a user performs their work.

This project is built upon the foundation of the biometric research already underway at AFIT (described in Chapter 3) and is also inspired by DARPA's Active Authentication project [21]. DARPA seeks to use biometrics to ease authentication, that is, to unobtrusively verify users during an active computer session. The initial phase of the project studies the 'cognitive fingerprints' [21] left behind by a user while interacting with the system — how words are crafted in documents or how the mouse is handled. This focus applies directly to Lt Bailey's research, conducted in tandem with this thesis and discussed in Chapter 3, and to the work done here in Chapter 6 analyzing documents.

III. Related Research

This chapter describes the research that is directly related to this thesis project, providing the project's foundation. It will go into a brief discussion of the computer based biometric research done in tandem with this thesis and conclude with the previous and ongoing biometric research involving gait analysis.

3.1 Computer Based Biometric Authentication

Kyle Bailey [22] has developed a tracking software to record the keyboard and mouse dynamics of a user interacting with a computer. The collected modalities are mouse clicks and movement, key presses, and GUI interaction. Features extracted from this data convey a user's habits in the use of a computer — for example, how long keys are held down, the average time between two key presses, and whether a user prefers the mouse to keyboard shortcuts.

Bailey used Weka, a data mining toolkit [23], to examine the features from each modality. Three sets of exemplars were generated from the tasks each user performed as described in Chapter V. A machine learning algorithm trained on two of these sets from each person as training sets. The third set was sent to the trained classifier, which tried to distinguish between the users. The Bayes Net algorithm proved to work best and was able to differentiate all the users. Additionally, Bailey found that fusing the modality features together yielded more accurate results than using a single modality on its own. Future work in this area will authenticate users 'live' while they are working, both for binary classification to make a yes/no determination if the current user has changed, and also for identification from a database [22].

Bailey's work focuses on authenticating users while they are actively pressing keys and using the mouse; in contrast, this thesis focuses on when users are briefly inactive.

Combining these two methods should result in robust and proactive authentication.

3.2 Biometric Gait Analysis

Anum Barki [24] [25] is investigating dismounted behavior concerned with the differentiation of individuals who may be carrying a load. This research is a continuation of Dr Kimberly Kendrick's [26] [27] analysis of the upper extremity during a gait cycle using Groebner Basis to solve the inverse kinematics problem. The inverse kinematics problem states that if the position and orientation of end point is known, though back substitution and the Groebner Basis, the angles of the joints can be found. In Kendrick's work, geometric equations were constructed, describing the geometry of the upper extremity system with 4 equations and 4 unknowns. These equations are too complicated to evaluate directly; but, by using the Groebner basis through the software Magma, simpler equations are produced which can then be solved for all possible solutions to the problem, including the no solution case.

Barki [25] has applied this work to the lower extremities, generating 6 equations and 6 unknowns for the 4 relevant joints - hip, knee, ankle, and base of toes. The solutions yielded by the Groebner basis will be applied to analysis of the leg behavior in the gait cycle phases when a loaded vest is and is not worn by a subject. The question here is whether a load causes the angles of the leg to change in a predictable manner while a person is walking [24]. Future work will compare this 2D model to a 3D model in development by Dr Kendricks.

Related work by Barki investigated the angles the back makes while walking up and downstairs while carrying a load, with initial results confirming the hypothesis — under a load the angle of the back will decrease in order to compensate.

IV. Phenomenology

In computer based work we want to tell the difference between experts and non-experts. Being able to identify expertise by a person's behavior — that is, their interaction with other people, objects, or information — lets us exploit opportunities and recognize threats. [28] [29]

A classic problem in distinguishing expert from pretender is authenticating the user of a computer. In this case, our expert is the authorized user, the person who most acts like the individual the computer is looking for.

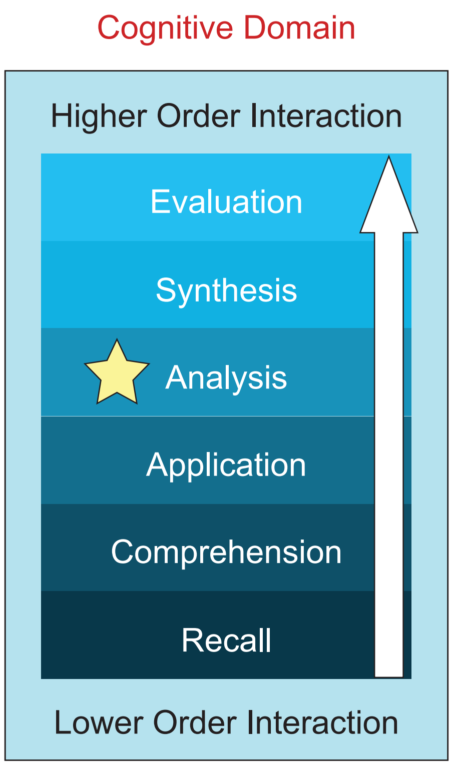
To proceed, what we require are methods that characterize behavior of users and an understanding on what behaviors to draw out and concentrate on. To ensure both uninterrupted access and vigilant computer authentication, we seek to identify that user whenever and however they return to work, even after others have used the computer. Therefore, in this chapter, we present the phenomenological basis for the focus our experimental design: "set" as a start/stop indicator of a task at an interface and "analysis" as an advanced information management task.

4.1 Bloom's Taxonomy

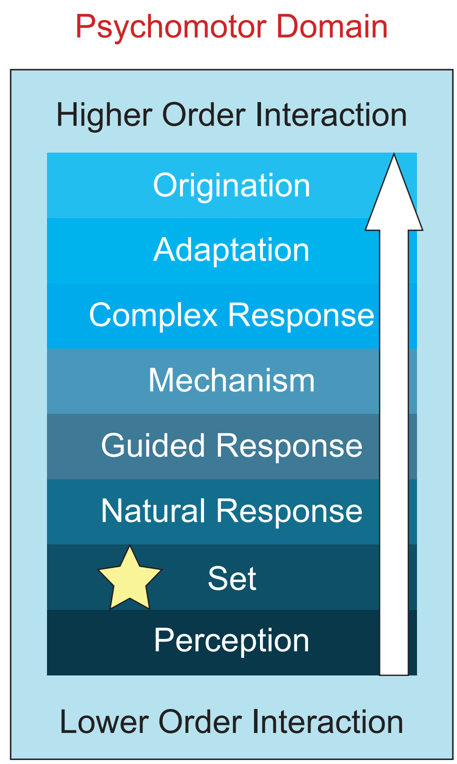
The phenomenology that guides this research is Bloom's Taxonomy. This theory was created by a committee led by Benjamin Bloom in 1956 and identifies three domains of learning: Cognitive, Affective, and Psychomotor. The Cognitive domain involves the development of abstract reasoning skills where a person interacts with information. The subcategories of the Cognitive domain are depicted in Figure 3 from simplest to most complex [30] are Recall, Comprehension, Application, Analysis, Synthesis, and Evaluation.

When evaluating proficiency, we look for verbs indicating application and analysis.

Verbs point to the thought processes behind sentence construction. A writing sample with application and analysis verbs is likely to be information dense and demonstrative of expert intelligence rather than that of a novice. By evaluating the verb lexicon employed to answer a query, and the timeliness and ease with which lexicon is employed, we reveal competency with the task in question. We seek to employ this analysis on the text documents produced by computer users during the experiment and correlate this higher level thinking to hand behavior patterns.



(a)



(b)

Figure 3. Two learning domains of Bloom's taxonomy: (a) Cognitive skills track interactions with information and (b) Psychomotor skills track interactions with interfaces. In this thesis, we hypothesize that we can use these taxonomies to assess skills in everyday work situations. In our experiments for this research, we focus on users as they display both set and analysis in the creation of a new document.

The Affective domain articulates interactions with other people [30] and is the least relevant to our current work. The Psychomotor domain [30] is the primary

domain that this research falls into. This domain involves a person interacting with the environment and its interfaces. This research seeks to evaluate the skill level of users by the method with which they interact with the computer.

The Psychomotor domain includes the following subcategories (Figure 3):

- Perception - awareness of surroundings [30].
- Set - readiness and intention to act the moment before an action occurs [30].
- Natural Response - initial action based on intuition, one's initial attempts at doing something. Note we propose adding this step to account for the coaching resistance. Before a subject can take guidance from a teacher, they must first develop a concept-to-action mapping by going through the action themselves.
- Guided Response - allowing a teacher to guide actions to form corrected behaviors, and includes imitation and trial and error [30].
- Mechanism - an action has become habitual and proficiency can be increased through targeted drills [30] [31].
- Complex Overt Response - mastering coordination of multiple skill sets [31].
- Adaptation - modification of skill sets to fit a changing environment [30] [31].
- Origination - creation of new patterns or skills in response to environmental demands [31].

A person in the Set level indicates a readiness to act — when using a keyboard, this indicates a readiness to begin typing a thought. This type of behavior — constructing what to say and then typing — occurs repeatedly while one types. Thus, we expect that the set position is a consistent behavior that will occur at a predictable rate.

In our experiment, we have the opportunity to combine the three modalities to characterize — (1) the video information on the subject's hand posture, (2) the verb analysis from the produced documents, and (3) event data from the keylogging — to construct a model that will evaluate a user's competency with a given task. Bloom's Taxonomy allows us to design an experiment based on predictable physiological behavior (i.e., the set position) to evaluate tasks (i.e., a cost benefit analysis). It also motivates us to assign free form tasks to computer users to ensure that they are not doing overly simple recall tasks but encouraging them to operate at higher levels. By challenging the user, we enable them to reveal their preferences, which can point out the uniqueness in a user.

V. Research Approach

5.1 Purpose and Objectives

Evaluating a user's competency establishes their credibility and authority, enabling recognition of potential persons of interest — experts, leaders, and potential threats. We seek to observe the posture of a computer user's hands as they transition between crafting ideas and writing them. These set positions are common postures, yet subtly different, among users. We will also test how well the set posture combines with keylogging data and text documents that the user composes in our assessment of a user's competency. We plan to evaluate a user's competency by the verb lexicon they use in relation to the task and subject, provided by the document they craft, and by the timeliness and ease with which they employ this lexicon, determined from the video and keylogging data.

This chapter describes the research approach to this end, starting with equipment and setup and the tasks through which participants produce the documents we will analyze. The chapter will also present the methods for identifying the set position, orienting of the video, and performing background subtraction and hand isolation. Finally we will discuss how user differentiation is performed and issues associated with the sensitivity and specificity of the hand model.

5.2 Equipment and Setup

Recordings of computer work took place in the Video Analysis and Context Extraction (VACE) Laboratory. The room is a standard indoor, climate controlled meeting room. It has dimensions of approximately 25 by 27 feet, and contains tables, chairs, computers, projection equipment, and white boards. We provided a standard DOD desktop connected to the Internet via DREN as the main station for the study.

The work area was furnished with a table and an adjustable chair and set similar to a cubicle but without the walls.

The room was equipped with the specialized recording equipment:

- Two DOD standard desktop computer with software for collecting behavioral biometric data specifically keypress, key release, mouse button presses, mouse scroll wheel movement at about 15 Hz, and a subset of GUI window interactions messages associated with user actions (clicks, resizes, text field, drop down, and radial button selections, etc.). The software also time stamps each data item.
- Two Creative Vado HD cameras to capture high resolution, high frame rate images of subjects hand position. The Vado HD cameras were positioned to capture the forearms and hands of the subjects.

Initially, the Kinect was selected for use in this experiment. The Microsoft Kinect is a popular camera because of its synchronization of depth and RGB data, and its availability and low cost. However, we found that when the hand is near an object or interacts with an object, such as typing on a keyboard, they are difficult to tell apart using the depth sensor. We attempted background subtraction methods using the Kinect's RGB data, but without success, and transitioned to smaller, more capable RGB video camera, the Vado HD camera. The Vado HD performs at 30 frames per second at 1280 by 720 pixel resolution, which is crucial for observing small, fast events, such as typing.

Each workstation was equipped with an internet-capable computer, keyboard, and mouse. Additionally, a Creative Vado HD camera was positioned overhead to record the typing hands of a user (See Figures 5 and 6). Users were allowed to change the computer configuration of the keyboard if desired - Microsoft Windows allows users to switch between QWERTY and DVORAK (Figure 4) layouts in the software.

Figure 4. DVORAK keyboard layout. [Public domain image].

Each keyboard was black, and a black cloth was placed underneath to cover the area in the camera frame of view where a user's hands were located while typing. We observed that a dark background yielded the greatest contrast between foreground (hands) and background (keyboard/desk), enabling more consistent hand tracking and reliable data collection. Participants were instructed to not move the keyboard during the typing session to ensure that we collected the hand pose completely and consistently during the session.

Video was captured using a Creative Vado HD camera at 720 x 1280 pixel resolution using H.264 compression. The Vado HD was chosen for its high frame rate (30 fps) and high definition video. The Vado HD was attached to a tripod via a horizontal metal rod, suspending the camera approximately 54 centimeters over the keyboard of a computer workstation, out of the way of potential users. Two setups were created to allow for data collection of two users at once.

The software Fiji ImageJ converted the Vado videos into frames for rotating and cropping. Because the Vado cameras save video as avi files with a codec not compatible with ImageJ, VirtualDub was first used to resave each video file as an avi with an ImageJ-compatible code before frame conversion.

Each computer had Bailey's software installed on it for key logging and mouse

Figure 5. Photograph of the laboratory setup for two users. Both workstations have an identical setup, aside from the monitors. The Vado HD camera is connected to a tripod 54 cm above the table via a metal rod, allowing the camera to record typing without disrupting the user. A black cloth is placed underneath the keyboard for better contrast between hand and background during background subtraction. Blue tape on the table delineates for the user the approximate frame of view of the Vado camera to help the user keep other objects outside of this field of view.

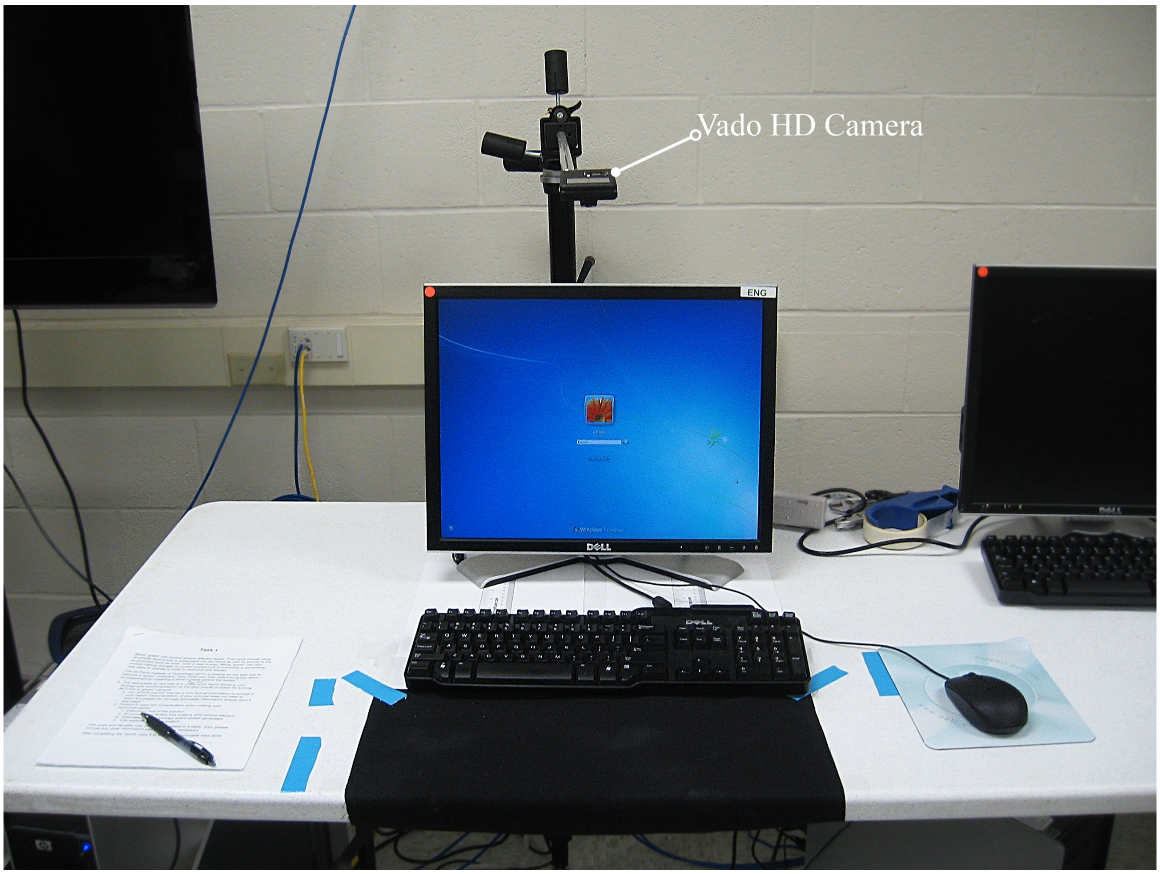


Figure 6. Workstation from user's point of view. To avoid disruption, the Vado HD camera is suspended above keyboard via the metal rod connected to a tripod, out of the user's field of view of the monitor.

click detection and recording. It ran in the background of each computer and silently recorded data while the user worked. Available computer applications for the user included the internet browsers Internet Explorer Version 9.0.8112.16421, Google Chrome Version 25.0.1364.152, and Firefox Version 15.0.1, and word processing applications in Microsoft Office 2010 Version 14.0.6129.5000.

5.3 Experiment Tasks

This experimental approach differs from other studies in that, rather than focusing on the action, we are focused on the pauses in movement — the set position, the transition from not typing to typing. Additionally, rather than using a rote task, we allow the typist to improvise, potentially using comprehension, application, analysis, and synthesis, exercising skills that are not just simple recall.

The experiment was conducted in several stages: 1) typing test, 2) background capture, and 3) green energy proposals. The mouse click and key logging software was used during the three tasks, in conjunction with Bailey's research work [22].

During Stage 1, participants were asked to take a short 500 character typing test to determine each participant's typing speed in words per minute. This typing test is located at: http://www.lecturel.com/clavier/words-per-minute.php.

Stage 2, background capture, occurred next. After starting the video recording, participants waited at least 40 seconds before beginning their work to allow the vibration from the interaction with the Vado camera to cease, and to allow the Vado camera to capture background frames of the keyboard and tabletop with no hands in the frames. The initial frames were used as the background during the background subtraction.

For Stage 3, participants were then instructed to type 400-500 words each on three topics involving different green energy solutions for AFIT. The topics were chosen

to incentivize participants to research a short proposal, and to encourage complex interaction with the computer both in terms of the mouse and keyboard interfaces and various computer applications. Any additional time remaining in the three hour study was given to the participant to type on work of their choice — suggestions included thesis, class reports, and dissertations, in order to examine computer based work that was more familiar to the participant and thus more demonstrable of the participant's expertise. Internet access was provided for research. Each typing session was approximately 2-3 hours, allowing the user enough time to become comfortable with the setup and type naturally on the given topics. The Vado camera recorded the participants' typing during the three tasks. Breaks were allowed if desired and did not count toward the time limit.

5.4 Video Analysis

Data analysis was broken into two sections, video analysis, and green energy proposal document analysis. Immediately following is a discussion on the video analysis, followed by a section on proposal analysis.

Video analysis was broken in to several stages: 1) Collecting Overhead Video and Identifying Set Position, 2) Processing Video, 3) Background Subtraction and Hand Isolation, 4) Ellipse Extraction, and 5) Participant Differentiation. These stages will be discussed in detail followed by some issues and limitations.

Collecting Overhead Video and Identifying Set Position

The set position is the precursor to action and, in typing, an easily defined position with respect to standard keyboards. Video for each participant was visually studied to identify frames where the user's hands were in the set position. In order to identify these set position frames for either hand, video for a participant was studied to

determine the position that the hands consistently returned to. Once found, the fingers were watched during their movement to find the frames when the fingers stopped moving from their keystroke and settled in to the set position, and also, to find the exact frames when the fingers started to deviate from their set position.

Processing Video

Video processing occurred in these steps: 1) Choose video sections, 2) Record 'Set' frames for database, 3) Crop and rotate frames.

In Stage 1, sections of video between twenty and forty seconds long were extracted from times corresponding to approximately near beginning, middle, and end of a participant's typing session. For each section, a frame with just the static background — keyboard, desk, and black cloth — was selected as the 'background frame' for background subtraction.

From the 20-40 second sections, in Stage 2, we recorded frame numbers corresponding to set positions of left and right hands in an Excel spreadsheet.

A consistent frame size was needed in order to compare properties dealing with the locations of the hands. Therefore, in Stage 3, using Adobe Photoshop to measure pixel locations in each video, all frames were rotated if not square (squareness was based on degree of rotation of the top edge of the keyboard to the horizontal) and cropped to the far right and top edges of the keyboard (as viewed from the perspective of a typist), and to approximately one inch below the base of the left thumb during the lowest portion of that hand's set position (see Figure 7). Frames were rotated generally between 0.08 and 2.00 degrees until square, with an error of ± 0.04 degrees.

Cropping and rotation created a consistent coordinate system with the origin at the top right of the keyboard. The x-axis progresses to the left of the keyboard when oriented normally to a user (to the right in the camera's point of view), and the y-axis

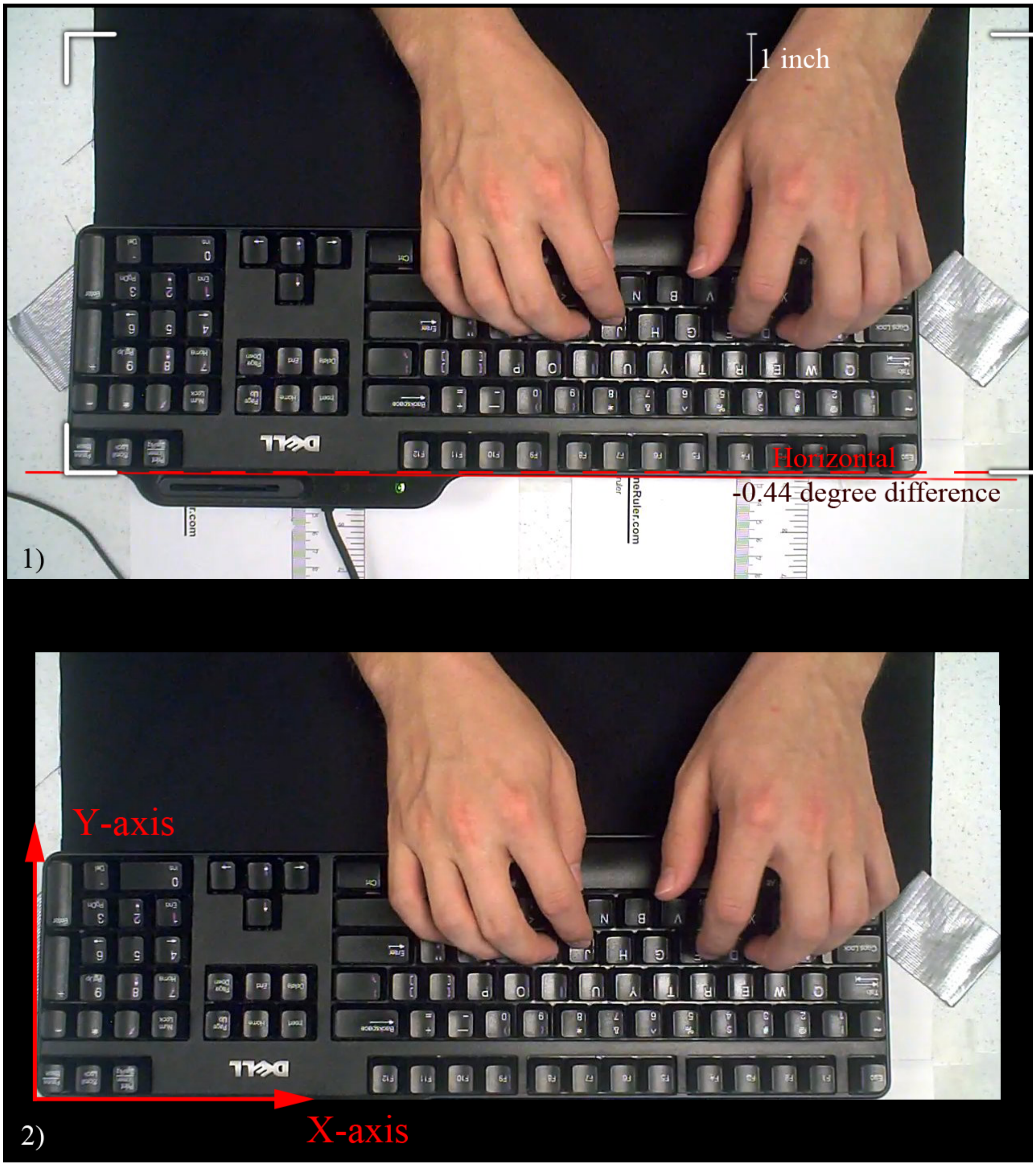


Figure 7. Rotation and cropping of video frames. Image 1 shows the raw video frame before processing. The frame angle is measured with respect to the horizontal and the top edge of the keyboard, then rotated appropriately to square the image. The 1 inch measusrement indicates where the top of the frame is cropped to with respect to the base of the left thumb in the set position. The frame is cropped to this measurement and to the top and right of the keyboard (viewed as a typist) once rotated. The white corners mark the boundary to which the image will be cropped. Image 2 shows the rotated and cropped image. The coordinate axis indicates the origin in the image, and the directions of the y- and x- axes. The black border is only to indicate the size of the reduced image and is not a part of the image frame during processing.

progresses from the top of the keyboard to the bottom. Since the x-axis progresses to the camera's right, the right side of the frames were left untouched — the *length* the frame does not matter as long as the origin is located at a consistent point.

The choice of cropping the video one inch below the base of the left thumb (see Figure 7) was made to ensure that 1) during the majority of typing, both hands are fully in the field of view, and, most importantly, 2) that each user is compared with consistent anatomy. Since the exact point where the base of the hand becomes the wrist is not always readily apparent in the video, a point was chosen — the base of the thumb — that was usually identifiable. Measuring one inch below the thumb's base ensures that the entire hand will be located in the analyzed frames. The left hand's thumb was chosen for consistency, and, by visual inspection of the videos, the left hand in the set position was usually lower than the right hand's set position. Therefore, measuring from the left hand allowed the greatest probability that both hands would be fully visible in the frames during their respective set positions.

The thumb base measurement was done for each participant. Therefore, each participant's frame height will not necessarily be the same due to hand anatomical differences. Participants with larger hands will naturally have larger measurements — for example, the length of the hand. Different frame heights highlight the anatomical differences and are not detrimental to the way the analysis is conducted — if instead, all frames were cropped to the same height, then, for example, when comparing the length of two different participants' hands located in the traditional typist 'home' position on the keyboard, the analysis could indicate that the hands are near the same length, when in fact, they could be completely different sizes. Analysis is done using background subtraction, which reveals the hands and any wrist in the frame, and an associated ellipse that best circles the foreground area, including the hand and wrist. For identically cropped frames of the home position, one of a large hand

with little wrist in the frame, and one of a small hand with lots of wrist in the frame, the encircling ellipse will have approximately the same length for each hand, nullifying any real comparison. We are not attempting automated image processing but a disciplined procedure for comparing consistent anatomy — therefore, a larger hand will naturally have a larger frame height than a smaller hand, and subsequently, a larger ellipse, making comparisons between different people possible.

The method of cropping the frames with respect to individual hand size does not bias the results of this experiment. This experiment considers not just pure behavior (invariant to scale) but also the physical features of the hand, i.e. size of the hand. Both are important in biometric authentication methods in distinguishing between individuals. In this particular method, distinguishing based on the set position, multiple typists may have a similar hand size, or a similar set position. It is the fusion of the data that enables the best differentiation, rather than one modality alone. This method of cropping enables the ellipse to capture the relative sizes of participants' hands for comparison, and therefore, the differentiation results are based on behavioral biometrics and physical biometrics.

Background Subtraction and Hand Isolation

As mentioned, background subtraction was used to isolate the hands. Using an identified background image frame, a Matlab algorithm subtracted this frame from each frame to be analyzed and colored white any pixel that was both skin colored and over a high contrast threshold. All other pixels were colored black. This algorithm created a binary black and white image, with the hands in white, which was then used for analysis (Figure 8). This process is illustrated here:

Each image $\overrightarrow{p}$ can be described as $\overrightarrow{p} = (\overrightarrow{p}_{11}, \ldots, \overrightarrow{p}_{ij})$, where i is the number of columns in the image and j is the number of rows, and a pixel $\overrightarrow{p}_{ij} = (r_{ij}, g_{ij}, b_{ij})$,

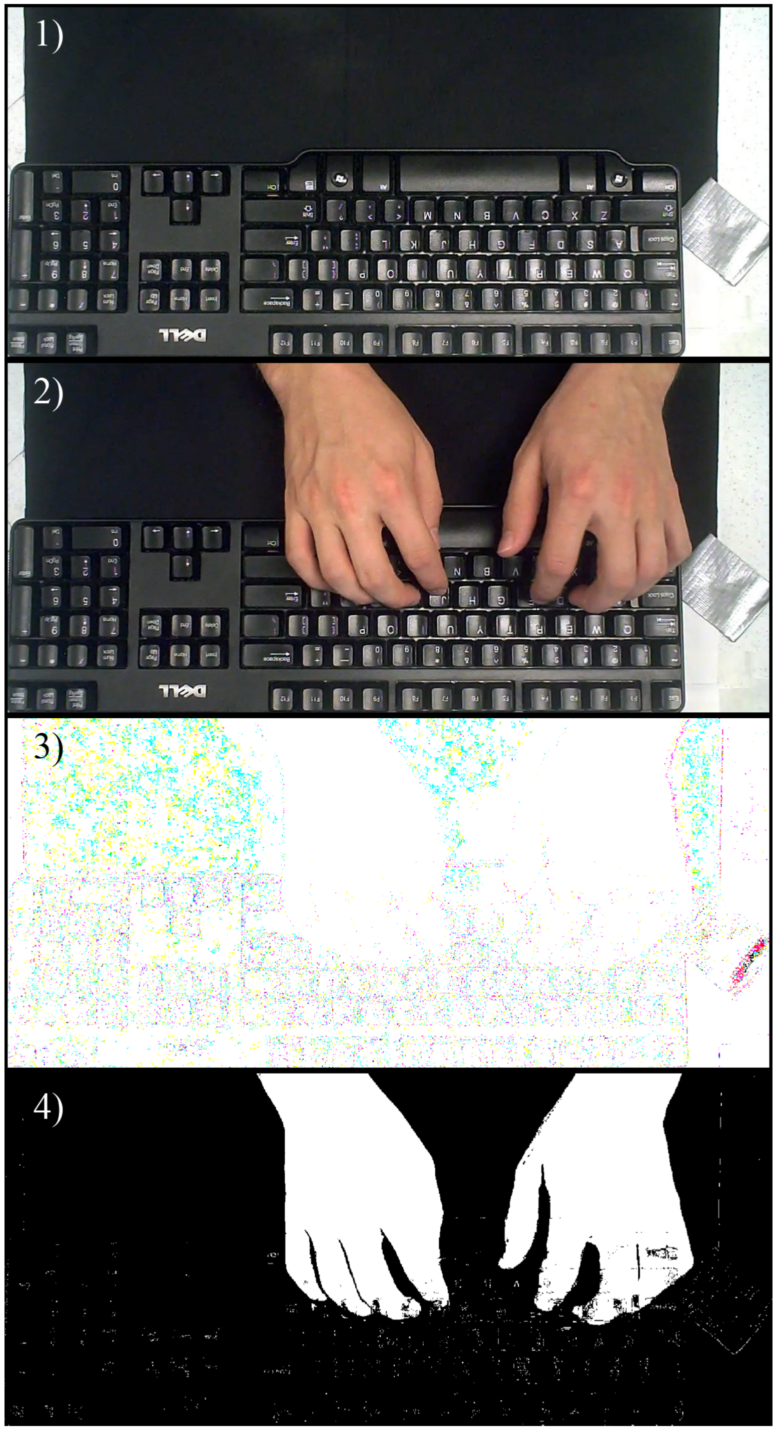


Figure 8. Background subtraction and hand isolation. Image 1 shows the cropped and rotated background image, with no hands in the frame. Image 2 shows an example frame to be analyzed. Image 3 shows the image after background subtraction, but before the binary image is produced. Image 4 shows the binary black and white image, where the hands are white and the background is black.

where r, g, and b denote the red, green, and blue channels. The subtracted image, where the static background image (no hands - see Figure 8 Image 1) is subtracted from the current image (with hands - Figure 8 Image 2) is

$$s = |\ \vec{p}_c - \ \vec{p}_{bg}|$$

where c denotes the current frame, and bg denotes the background frame. The rough subtracted image is shown in Figure 8 Image 3.

In order to further isolate the hands, we determined that, because of the skin color, there was a high contrast between the hands and the mostly black background, as seen in Figure 8 Image 2. This sharp distinction of the hands from background enabled the use of a contrast threshold which, after background subtraction, tests the remaining red, green, and blue pixel values. Values had to be much greater than — or much less than — threshold to be considered part of the hand:

$$|\ \vec{p}_c - \ \vec{p}_{bg}|_r > \rho$$

$$|\ \vec{p}_c - \ \vec{p}_{bg}|_g > \gamma$$

$$|\ \vec{p}_c - \ \vec{p}_{bg}|_b > \beta$$

where r, g, and b, again denote the red, green, and blue channels, ρ is the threshold for the red channel, γ is the threshold for the green channel, and β is the threshold for the blue channel. We determined experimentally that values that worked well for these three thresholds were 30, 60, 60, respectively.

To add even greater ability to isolate the hands, the skin hue of the current frame c (Figure 8 Image 2) was used for another set of thresholds:

$$\overrightarrow{p}_{cr} > \rho_c$$

$$\overrightarrow{p}_{cg} > \gamma_c$$

$$\overrightarrow{p}_{cb} > \beta_{c,low}$$

$$\overrightarrow{p}_{cb} < \beta_{c,high}$$

where $\overrightarrow{p}_{cr}$, $\overrightarrow{p}_{cg}$, and $\overrightarrow{p}_{cb}$ denote, respectively, the red, green, and blue channels for the current frame. Values that were experimentally determined to work well for ρ_c, γ_c, $\beta_{c,low}$, and $\beta_{c,high}$ for most participants were 106, 67, 60, 170. Adjusting these numbers on a case by case basis would result in better isolation of the hands for that individual. In the future, a learning algorithm should be employed to achieve better individual results.

Absolute value was used in these calculations for a cleaner background subtraction. Figure 9 shows the slight difference between the binary image without using absolute value (top) and using absolute value (bottom).

The pseudocode that describes the hand isolation process is as follows:

for ALL pixels **do**

if $(\overrightarrow{p}_{cr} > \rho_c$ ***and*** $\overrightarrow{p}_{cg} > \gamma_c$ ***and*** $\overrightarrow{p}_{cb} > \beta_{c,low}$ ***and*** $\overrightarrow{p}_{cb} < \beta_{c,high})$ ***and***

$(|\overrightarrow{p}_c - \overrightarrow{p}_b|_r > \rho_s \; AND \; |\overrightarrow{p}_c - \overrightarrow{p}_b|_g > \gamma_s \; AND \; |\overrightarrow{p}_c - \overrightarrow{p}_b|_b > \beta_s)$

then

set current pixel to white

else

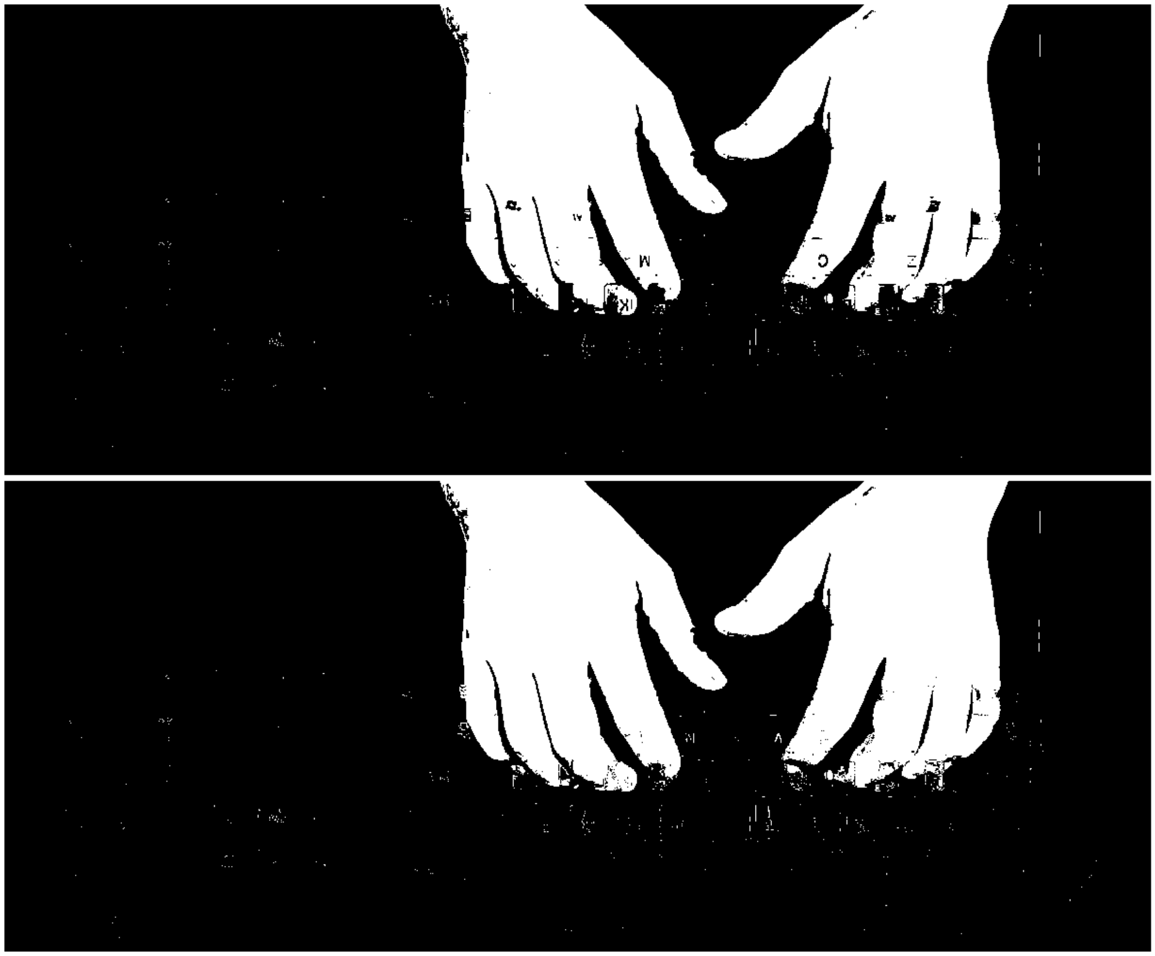


Figure 9. Comparison between binary images when using absolute value. Top: binary image without using absolute value during calculations. Bottom: binary image using absolute value during calculations.

set current pixel to black

end if

end for

After the binary black and white image was created, we looked at an ordering (Figure 10) of connected component size to number of connected components. The hands were readily identified as the only two objects over 30,000 pixels in size — this number was used to isolate them from background noise left over from the image processing.

The Matlab function *regionprops* measures properties for connected regions and achieves segmentation and feature selection. Its property, *area*, which contains the sizes of each connected region, was used to extract the two hand regions identified by the ordering. *Regionprops* was also used to extract 6 ellipse properties from the connected regions of the two hands. These properties are associated with ellipses that have the same second central moments of the two hand regions, and include 1) *orientation* in degrees, 2) *eccentricity* where a value of 0 specifies a circle and 1 a line segment, 3) *major axis length* and 4) *minor axis length* in pixels. Also extracted from *regionprops* was the *centroid* in pixels that specified the 5) x- and 6) y-coordinates of the center of mass of the region. These six properties defined an ellipse describing the hands' basic shape and position.

Ellipse Extraction

These ellipse properties were first extracted from both hands in frames containing their identified set positions. These hand selected frames created a database for each participant of set position ellipses. From this database, the maximum and minimum values of each property were used to create a maximum and minimum set position ellipse for each user.

Once these maximum and minimum set position ellipses were obtained for a user, larger video segments from that user were analyzed to obtain ellipses for both hands in all the frames, regardless of hand position. This data was then compared to the maximum and minimum set position ellipses — if a given hand ellipse fell in between the defined range, it was labeled as being in the set position for that user. This labeling was done for both the left and right hands, creating a file of all set position frames and ellipses for a user.

Regionprops labels the connected regions as they appear from left to right across

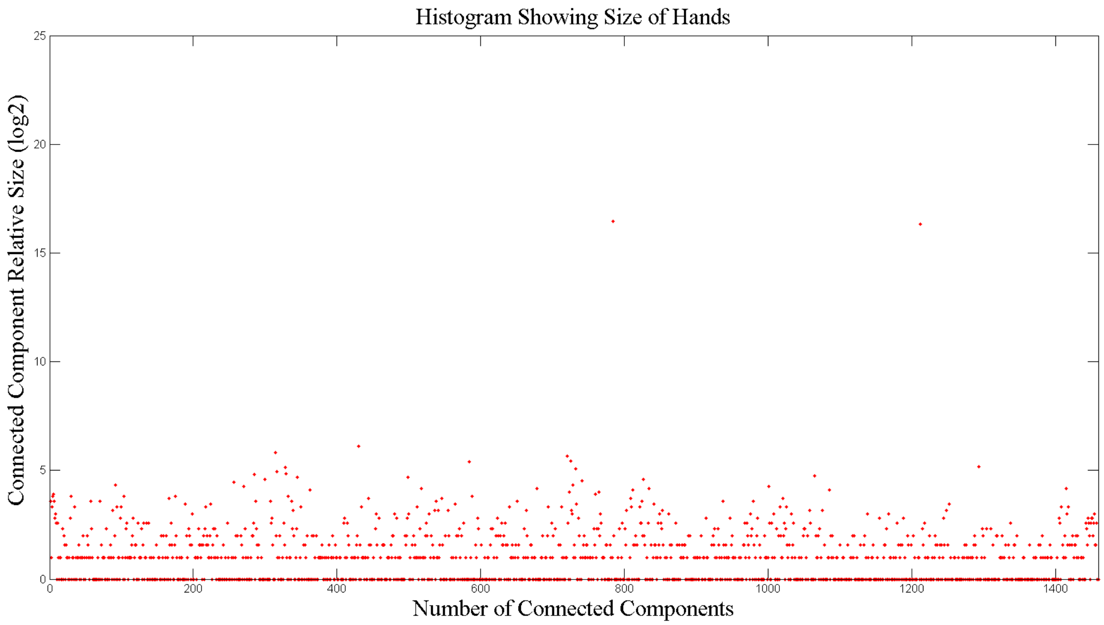


Figure 10. Histogram showing relative connected component size to number of connected components. The two largest data points indicated the large size of the hands compared to the rest of the speckle in the image.

an image. Labeling confusion can occur in a case where the left hand is in an image, labeled as Object 1, and then the right hand enters the image, causing the left hand to be relabeled as Object 2 while the right hand takes the label Object 1. Such a case occurs when the participant might be using the mouse with the right hand, leaving the left hand on the keyboard, before resuming typing with both hands. Therefore, in order to avoid this confusion, ellipse processing was only done when both hands were located within the frame (Figure 11).

Differentiating Between Participants

Once a database of set positions was created for each participant and after the larger video segments were analyzed to obtain a stream of ellipses from all the frames, the set positions were used to attempt differentiation between participants.

The ellipse data from the larger video segments of 10 participants was combined

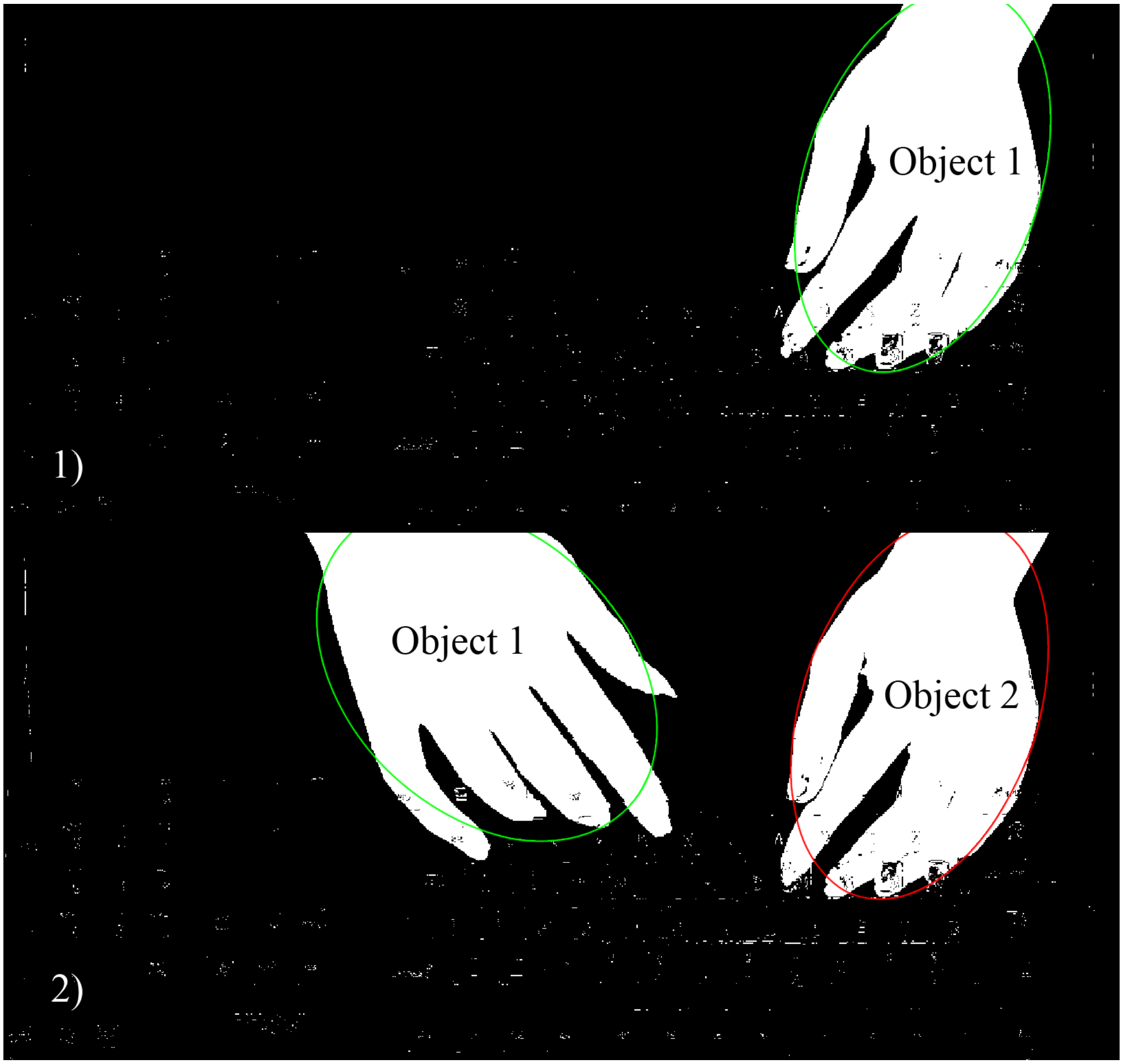


Figure 11. Labeling confusion. Image 1 shows how Matlab's *regionprops* function will label a single connected region, in this case, the left hand. Image 2 illustrates that *regionprops* labels regions from left to right across an image, causing labeling confusion where first the left hand was labeled as Object 1, and now it is labeled Object 2 as the right hand moves into the frame, taking the label Object 1. For this reason, ellipse processing was done only when both hands were in the frame to avoid labeling confusion. Also visible in both images are the ellipses conforming to the hand shape, created using the six ellipse properties in *regionprops*.

into one larger matrix. This matrix represented video from each participant in turn, essentially creating a scenario where the active user at the keyboard changes multiple times, illustrating our change detection scenario.

Each recorded ellipse (essentially each frame of each video) was compared to each user's set position range. If the ellipse fell within that range, the frame was identified as that user. This method sought to identify users based upon their set position only. Therefore, frames that did not have set positions in them were ignored.

Issues and Limitations

During the analysis of a participant's video segment to find frames that fell into the range of that user's set position, several issues were identified. Although frames visually identified for the user's database were indeed set position frames for either the left or right hand, a person's set position may have some wide variation in one or more of the six ellipse properties recorded. This variation, in turn, may produce false positives (that is, the participant may actually be striking a key or moving to strike a key) when the identification code is run on the entire video segment. False positives may be especially troublesome for participants whose set position is located on the home row, where a traditional typist's set position would be located. When a participant strikes keys on the home row, namely those underneath the index through pinky fingers — A, S, D, F, J, K, and L — differentiating those ellipse positions from the set position is difficult using the ellipse properties since the hand does not need to move much during those events. The ellipse of a hand striking any of those keys may be mislabeled as a set position. The same may hold true for keys on the bottom row — Z, X, C, M, and possibly V or N — depending on the amount of variance that was recorded for the participant's database of set positions.

This issue may hold especially true for users of the DVORAK keyboard layout,

where approximately 70% of keyboard strokes are done on the home row.

Issues were also identified when using the set position to differentiate between participants. When only a few users are compared, we can to tell them apart with relative ease, but, as the pool of users enlarges, we will more and more likely find different users who have similar set positions, that is, because of the variance in users' set positions, an ellipse that describes one user's position (say, User 1) at one point in time in a frame (not necessarily in User 1's set position) may fall inside the range of User 2's set position, and thereby be labeled as User 2.

If the ellipse in question also happens to describe User 1's set position in a frame, then this ellipse may be labeled as both User 1 and User 2, resulting in a 'confused' detection.

We might be able to correctly identify a questionable detection by looking at the frames surrounding it. If the majority are labeled as User 1, we can reasonably assume that the frame in question is also User 1 as opposed to User 2, especially if it is just a single frame labeled as User 2. With a camera running at 30 fps, User 2 is not likely to physically show up for just one frame.

5.5 Proposal Text Analysis

We began analyzing the documents produced over the course of the experiment with regards to Bloom's Taxonomy. Because verbs point to the thoughts behind language, initial analysis consisted of identifying the verb phrases used. Each paragraph produced in a document was then ordered by its level, keeping in mind Bloom's levels of the Cognitive domain: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. Paragraphs employing Application and Analysis were marked as 'high level', while paragraphs employing Knowledge or Comprehension were marked as 'low' or 'medium'. Figure 12 shows this initial analysis, done by Dr Magnus. The

top half shows the main verbs phrases for each sentence in each paragraph. The bottom half shows all the verb phrases in addition to the relative level of each paragraph. Results gained by fusing this work with the video will be described in Chapter 6.

	A	B	C	D	E	F
1	**Subject 6 Main Verbs**	**Paragraph 1**	**Paragraph 2**	**Paragraph 3**	**Paragraph 4**	**Paragraph 5**
2	**Sentence 1**	does	require	are projected	seems	is
3	**Sentence 2**	is	is	estimate	does not provide	are sited
4	**Sentence 3**	generate	is	make take	will mitigate	does have
5	**Sentence 4**	can be used		can save	shift	seems
6	**Sentence 5**	are		is	is	is
7	**Sentence 6**	don't take		is used		may be found
8	**Sentence 7**			appears		
9	**Sentence 8**					
10	**Subject 6 All Verbs**	**Paragraph 1**	**Paragraph 2**	**Paragraph 3**	**Paragraph 4**	**Paragraph 5**
11		**Medium**	**Medium**	**High**	**Medium**	**Low**
12	**Sentence 1**	does; recycle	require; designated, absorb, store, heating	are projected, using	seems; increase, become	is;uses
13	**Sentence 2**	is; is; see, exist	is; encounter	estimate	does not provide	are sited
14	**Sentence 3**	can be; generate	is; are required, to circulate	consider; make take;	will mitigate	does have
15	**Sentence 4**	can be used		can save	will shift; is	seems
16	**Sentence 5**	are		is	is; how much does?, use	will be; answering, is
17	**Sentence 6**	don't take		is used		may be found; investing, may be
18	**Sentence 7**			appears; provided, is		
19						

Figure 12. Example of Verb Analysis for User 6. The top half shows the main verbs phrases for each sentence in each paragraph. The bottom half shows all the verb phrases in addition to the relative level of each paragraph. Analysis was conducted by Dr Magnus.

VI. Results and Discussion

This section presents the results obtained from differentiation and the combination of the modalities used in this thesis, as well as some limitations.

6.1 Differentiation Between People

In a pool of ten participants, individual set positions may be readily discovered that distinguish between the participants. The six ellipse properties — centroid x-coordinate, centroid y-coordinate, orientation, eccentricity, major axis, and minor axis — were used as input features. Figure 13 shows an example of input features collected, a small subset from the left hand of User 6. The ‘maximum’ ellipse was defined from the 6 maximum property values; likewise, the ‘minimum’ ellipse was defined from the 6 minimum property values.

A random hand ellipse extracted from a video frame was labeled as a user if that ellipse’s six properties fell between the maximum and minimum ellipses defined in the user’s database. Excel was used to calculate statistics on the users detected during the experiment. Figure 14 shows an example of the data calculation for the left hand of User 6. As described in Chapter 5, data from video segment frames for each user were strung together in one large matrix, then read in by Matlab to attempt identification. All frames were renumbered, starting from 1. Column 1 is simply the frame designator that a detection occurred in and can be traced back to the actual image, but does not correspond to the actual frame number from the video. Column 2 is the ID number of the user detected in the given frame. Column 3 shows the number of times a user’s set position was detected for a given user, (in this example, User 6). The Excel formula Countif was used for this calculation. In the matrix mentioned above, User 6 was typing for frames 1-570.

Frame	X-coord	Y-coord	Orientation	Eccentricity	Maj Axis	Min Axis
19	973.6287	119.4343	-24.4195	0.5586	169.0737	140.2383
20	943.1552	143.1602	-61.8915	0.404	169.7807	155.312
21	925.7024	159.6986	-87.8548	0.5061	178.9629	154.3557
22	919.6047	168.9952	86.2954	0.5958	184.7814	148.4015
23	925.9199	173.4558	86.7823	0.6378	185.6214	142.9733
24	933.9974	170.791	85.8156	0.6329	181.2615	140.3436
25	937.6227	165.7306	86.5420	0.607	176.0019	139.8734
26	937.4357	158.2603	85.3166	0.5975	172.0003	137.9201
27	936.4805	151.4865	85.8262	0.5732	166.758	136.6393
28	933.0467	148.8987	82.5323	0.5581	165.3286	137.1892
29	930.5946	146.1365	81.7991	0.5489	163.3981	136.5814
30	927.1662	146.6246	80.9393	0.568	164.1509	135.1003
31	922.6305	147.667	77.9453	0.586	165.2969	133.9429
32	917.7903	149.8871	74.7507	0.5984	166.7542	133.6082
33	914.7639	151.8236	74.5974	0.6376	169.7086	130.7355
34	914.6296	153.2746	73.9746	0.6594	171.4584	128.8986
35	917.8257	153.0895	72.4374	0.6613	170.6268	127.9958
36	920.3278	153.1738	71.0702	0.6692	171.1612	127.181

Figure 13. A sample of hand data collected — that is, the elliptical statistics collected from a user's hand shape. Shown are sequential frames and the associated ellipse data — these are not just apparent set positions, but all sequential ellipses in a video sample. This specific data sample is from User 6, left hand.

180	442	6				
181	446	6				
182	447	6	#6 Indices	Left Hand		
183	456	6	1-570	Total		
184	457	6	Number of Counts	Detections	Percentage	
185	483	6	185	185	100	#times User 6 was detected
186	601	8	0	185	0	#times User 8 was wrongly detected while User 6 was typing
187	602	8	COUNTIF(B1:B185,9)	185	(D187/E187)*100	# of User 9 was wrongly detected while User 6 was typing
188	603	8	COUNTIF(B1:B185,10)	185	(D188/E188)*100	#times User 10 was wrongly detected while User 6 was typing
189	604	8	COUNTIF(B1:B185,13)	185	(D189/E189)*100	#times User 13 was wrongly detected while User 6 was typing
190	605	8	0	185	0	#times User 14 was wrongly detected while User 6 was typing
191	615	8	0	185	0	#times User 15 was wrongly detected while User 6 was typing
192	616	8	0	185	0	#times User 16 was wrongly detected while User 6 was typing
193	617	8	0	185	0	#times User 17 was wrongly detected while User 6 was typing
194	618	8	0	185	0	#times User 23 was wrongly detected while User 6 was typing
195	619	8				
196	620	8				

Figure 14. Example of Microsoft Excel work. This specific example is from User 6, left hand. The data to the left of the computations respresent a subset of the User 6 results.

Column 4 shows the total number of set positions detected for all users during the time the given user was typing. Column 5 calculates the percentage. Column 6 shows the user calculated for each row. In Figure 14, Column 2 shows the transition from the section of frames where User 6 was typing to the section where User 8 was typing. The calculations shown here are only on the segment of User 6, and end before User 8. The transition between Users 6 and 8 was included to show an example of the results obtained.

When taking into account the total number of set positions detected from all users and the total number of positions that were only labeled as the correct users, the accuracy was 92% in the video sections analyzed. Taken separately, the accuracy was 91% for the left hand (See Table 1) and 93% for the right hand (See Table 2). Out of 1730 set positions detected for the left hand, a total of 154 were labeled as an incorrect person. This number included detections that were labeled as both an incorrect person and as the correct person at the same time, indicating confusion in the set positions between those people. Of all these set positions that were incorrectly labeled, 42 were tagged as the aforementioned confused detections. The rest were tagged as set positions of users other than the correct user. In these frames, the correct user was not actually in his set position, but the ellipse describing his hand at that moment was similar enough to another person's set position range to be labeled as that other person's set position.

Table 1. Total Set Positions, Left Hand

Left Total Detections	Correct Detections with Confusion	% of Correct	Incorrect Detections	% of Incorrect	
1730	1618	93.526%	112	6.474%	
Left Total Detections	Correct User Only	% of Correct	Incorrect and Confused Detections	% of Incorrect	Confused detections
1730	1576	91.098%	154	8.902%	42

Out of a total of 1737 set positions detected for the right hand, 121 were labeled as an incorrect person (See Table 2). Of those 121 detections, 43 were confused detections where the correct user was also labeled as an incorrect user.

Table 2. Total Set Positions, Right Hand

Right Total Detections	Correct Detections with Confusion	% of Correct	Incorrect Detections	% of Incorrect	
1737	1659	95.509%	78	4.491%	
Right Total Detections	Correct User Only	% of Correct	Incorrect and Confused Detections	% of Incorrect	Confused Detections
1737	1617	93.092%	121	6.966%	43

The ten participants analyzed were users 6, 8, 9, 10, 13, 14, 15, 16, 17, and 23. The error matrices shown in Tables 3 and 4 break down the set positions detected while each respective computer user was typing. Table 3 shows the detections for the left hand. The typists are listed along the left, and the body of the table denotes the number of times that a user other than the current typist was incorrectly identified. Table 4 shows the same results for the right hand.

6.2 Labeling Errors: unique or confused set position detection

What these tables don't illustrate are the different cases when one user is mistakenly identified as another user. There are two types of this mistaken identification or mislabeling: 1) 'confused detections', cases in which the ellipse describing the current hand position is labeled as both the current typist and as an incorrect typist (when the current ellipse falls into more than one typist's set position range), thereby causing confusion as to the proper identification, and 2) 'unique detections', cases in which the ellipse describing the current hand position is labeled as only an incorrect typist. These 'unique detections' are likely cases where the current typist is not in a

set position, but the ellipse describing the position of the hand at the time falls into an incorrect typist's set position range. Confused detections and unique detections are listed in Appendix A.

6.3 Individual Results

User 6, User 13, and User 23, the only user who typed with a DVORAK keyboard, had a 100% of set position identification accuracy for both left and right hands. All other users had labeling errors which will be discussed further in this section.

User 8 had the lowest overall set position identification rate while they were typing. For the left hand, 15 detections were uniquely labeled as User 14, thereby resulting, out of 100 total separate detections (85 for User 8 and 15 for User 14), in an identification accuracy of 85%. For the right hand, 86 detections were labeled as User 9. Of these 86 detections, 31 were confused detections with User 8, and 55 were unique detections of User 9. These 31 confused detections are also counted among the 82 detections of User 8. Therefore, out of 137 separate detections (82 for User 8 and 55 unique detections for User 9) the identification accuracy for the right hand is approximately 59.9%.

While User 9 was typing, for the left hand, 9 detections were uniquely labeled as User 8, and 10 were labeled as User 10 (2 confused and 8 unique). An identification accuracy out of 165 separate detections for the left hand (148 for User 9, 9 unique for User 8, and 8 unique for User 10) was 89.7%. For the right hand, 1 detection was labeled as User 8 (confused) and 16 were labeled as User 10 (11 confused). An identification accuracy out of 179 separate detections for the right hand (174 for User 9, and 5 unique for User 10) was 97.2%.

While User 10 was typing, for the left hand, 33 detections were labeled as User 8 (15 confused and 18 unique), 31 detections were labeled as User 9 (25 confused

Table 3. Comparison of Detections Among Users for Left Hand

	# Times User Detected During Respective Typists												
Left Hand	6	8	9	10	13	14	15	16	17	23	# True Neg	# False Pos	Specificity
Current Typist: 6	185	0	0	0	0	0	0	0	0	0	1665	0	0.989356
8	0	85	0	0	0	15	0	0	0	0	750	15	
9	0	9	148	10	0	0	0	0	0	0	1313	19	
10	0	33	31	271	0	0	0	0	4	0	2371	68	
13	0	0	0	0	138	0	0	0	0	0	1242	0	
14	0	0	0	0	0	171	0	0	0	0	1539	0	
15	0	0	0	0	0	2	183	0	0	0	1645	2	
16	0	0	0	0	0	0	0	112	0	0	1008	0	
17	0	48	0	3	0	0	0	0	112	0	957	51	
23	0	0	0	0	0	0	0	0	0	213	1917	0	

Table 4. Comparison of Detections Among Users for Right Hand

	# Times User Detected During Respective Typists												
Right Hand	6	8	9	10	13	14	15	16	17	23	# True Neg	# False Pos	Specificity
Current Typist: 6	202	0	0	0	0	0	0	0	0	0	1818	0	0.991896
8	0	82	86	0	0	0	0	0	0	0	652	86	
9	0	1	174	16	0	0	0	0	0	0	1549	17	
10	0	1	7	209	0	1	0	0	1	0	1871	10	
13	0	0	0	0	175	0	0	0	0	0	1575	0	
14	0	0	0	0	7	84	0	0	0	0	749	7	
15	0	0	0	0	0	0	226	0	0	0	2034	0	
16	0	0	0	0	0	0	1	266	0	0	2393	1	
17	0	0	0	0	0	0	0	0	158	0	1422	0	
23	0	0	0	0	0	0	0	0	0	83	747	0	

and 6 unique), and 4 detections were uniquely labeled as User 17. The identification accuracy out of 299 separate detections for the left hand (271 for User 10, 18 unique for User 8, 6 unique for User 9, and 4 unique for User 17) was 90.7%. For the right hand, 1 detection was uniquely labeled as User 8, 7 were uniquely labeled as User 9, 1 was uniquely labeled as User 14, and 1 was uniquely labeled as User 17. The identification accuracy out of 219 separate detections for the right hand (209 for User 10, 1 unique for User 8, 7 unique for User 9, 1 unique for User 14, and 1 unique for User 17) was 95.4%.

User 14 had a 100% identification accuracy for the left hand. For the right hand, 7 detections were uniquely labeled as User 13. The identification accuracy out of 91 separate detections for the right hand (84 for User 14 and 7 unique for User 13) was 92.3%

While User 15 was typing, for the left hand, 2 detections were uniquely labeled as User 14, for an identification accuracy out of 185 separate detections (183 for User 15 and 2 unique for User 14) of 98.9%. The right hand had a 100% identification accuracy.

User 16 had an identification accuracy for the left hand of 100%. For the right hand, 1 detection was uniquely labeled as User 15, for an identification accuracy out of 267 separate detections (266 for User 16 and 1 unique for User 15) of 99.6%.

While User 17 was typing, for the left hand, 48 detections were uniquely labeled as User 8, and 3 detections were uniquely labeled as User 10. The identification accuracy out of 163 separate detections (112 for User 17, 48 unique for User 8, and 3 unique for User 10), was 68.7%. The right hand had an identification accuracy of 100%.

6.4 Sensitivity and Specificity Analysis

Keeping in mind the criteria for identification is whether or not a given ellipse falls within a user's database, all values along the diagonal in Tables 3 and 4 are regarded as true positives, while any other values are regarded as false positives. Zeros not along the diagonal can be regarded as true negatives. The identification process does not reveal false negatives — cases where a typist is in a set position but is not identified as such — though use of the keylogging data might help to resolve such instances though not perfectly. The true/false positive and negative values are in reference to the identification code and do not take into account cases where (1) a typist can be visually seen to be in a set position but that position did not make it into the database, or (2) cases where a hand model is overly broad and picks up instances when a user is still typing.

Based on these true/false positive and negative definitions, we selected a generous discrimination criteria that in a narrow sense ensured 100% sensitivity — that is, the criteria picked up all posture instances where the ellipse that defines that posture falls into a user's database. We do not expect that the sensitivity is truly 100% given the incompleteness of database, but these initial results do suggest that a reasonably tight set of features can resolve an individual's set position. We can say more about the specificity of the feature set, and the results there are promising.

Specificity is defined as follows:

$$specificity = \frac{number\ of\ true\ negatives}{number\ of\ true\ negatives + number\ of\ false\ positives}$$

Each set position for a given typist where any incorrect typist is not mislabeled can be regarded as a true negative. For example, in Table 3 where User 6 has 185 set positions identified, User 8 has no mislabeling for each of those 185 set positions,

therefore having 185 true negatives. The True Negative column in Table 3 is therefore the product of the true positives and the number of typists (9) other than the current typist subtracted by the number of confused detections for each other typist. The False Positive column is the sum of all mislabeled detections during a typist's session. Confused detections and unique detections are listed in Appendix A.

6.5 Results of Higher Level Work Analysis

Here we analyzed paragraphs crafted by User 6. When analyzing the paragraph deemed the most interesting from Task 3 based on the verb content — paragraph 3 — the identification rate was about 98.943% overall, 99.5% for the left hand, and 98.4% for the right (See Table 5). There was only 1 detection labeled as User 8's set position for the left hand out of 197 set positions detected. For the right hand, 12 detections out of 1058 were labeled as User 9's set positions and 5 were labeled as User 10's. There were no confused detections for either hand. These results show that we can identify a user based on their set position when they are doing their most critical work, but there is a sensitivity issue that we must address in the left hand model. We resolved this issue by fusing the hand model features with the keylogging data and determining which features fell outside the left hand model of the set position. We will discuss the fusion process next.

Three Modality Fusion

Next we examine the results of our higher level work analysis to explore discrepancies between left hand and right hand results. These discrepancies are best looked at by examining the elliptical features individually for the left hand when synchronized with keylogging data.

We could not synchronize the Vado HD camera with the keylogging software

Table 5. Accuracy of Detection for User 6 During Task 3, Paragraph 3

Current Typist: User 6, Task 3, Paragraph 3			
Labeled User	Number of Detections	Number of Confused Detections	Number of Unique Detections
Left Hand: 6	196		
8	1	0	1
9	0		
10	0		
13	0		
14	0		
15	0		
16	0		
17	0		
23	0		
Total Detections:	197		
Right Hand: 6	1041		
8	0		
9	12	0	12
10	5	0	5
13	0		
14	0		
15	0		
16	0		
17	0		
23	0		
Total Detections:	1058		

during operation as the Vado camera only allowed file transfer mode when connected to a computer. Therefore, synchronization of video and keylogging data had to be done after the fact. Neither the software nor the camera recorded current date and time. The software recorded computer system time since boot up and time in seconds since the software activation. The video recorded time since the start of recording and frame count. Therefore, in order to synchronize the data from the keylogging software — the keystroke data — and the data from the videos, we reviewed the keylogging data to identify the first several keystrokes recorded in the keylogging software. Once those keystrokes were known, we reviewed the relevant video to visually identify the frame numbers during those first key presses. Since the camera recorded at 30 frames per second, we could match the frames with the proper keystrokes and seconds from the keylogging software.

The difficulty in this synchronization method lies in the fact that because the camera records at 30 frames per second, several frames are recorded during the short time span when a finger hits a key. The frame that correspond to the actual register of the keystroke by the keylogging software is uncertain. The synchronization uncertainty was found to be $\pm$ 1 or 2 frames. Checking several keystrokes may reduce this uncertainty, but, towards the end of a synchronized file (approximately 33,000 to 100,000 frames), it had accumulated to about $\pm$ 3–4 frames.

An example of the synchronized data is shown in Figure 15. The graph shows the trend in the X coordinate of the ellipse centroid for the right hand for User 6 during Task 3 at the beginning of Paragraph 3. The entire typing of Paragraph 3 occurred in about 8,000 frames. A single graph can not clearly depict the entire trend because of the data density, so Figure 15 shows about 500 frames, which is about 16.67 seconds of video. The keystrokes are shown along the top of the graph, rotated vertically for a better fit. Figure 16 shows an enlargement of a section of the

graph annotating the keylogs and an adjacent set event. Red data points designate instances where the ellipse for this hand was labeled as a set position for User 6. The three green horizontal lines, from top to bottom, are the maximum, mean, and minimum values for the right hand set position for this ellipse property. These values were determined from the set position database formed for each user. Let use note that no frames from Task 3, Paragraph 3 were used during the formation of this database. Remember that the set position is determined from the aforementioned six ellipse properties as a whole, and that even though much of the data occurs between the maximum and minimum for the x coordinate, an ellipse will only be flagged as a set position if all six properties agree.

During the time frame that this graph covers, we observe that there are periods where the user does not type. With a traditional typist, one would expect the hands to remain in the 'home' position, which would also equate to that typist's set position. A set event on the left hand is in fact what is occurring between approximately frames 300 and 375 in Figure 15, and the right hand is identified as being in the set position. However, in the same graph of x coordinate vs time for the left hand (Figure 17), the left hand is not identified here, and this discrepancy between the hands points to an error in the set position database for the left hand, where perhaps the range of one or more other properties was defined too narrowly.

A review of the other left hand ellipse properties (See Figures 18 and 19) for this section of video reveals that only the orientation of the left hand ellipse was defined too narrowly. Figure 20 illustrates where the values are just above the defined maximum orientation for the frames in question, 300–375.

The overly narrow definition of the left's orientation feature explains why there were comparatively few left hand set positions identified compared to right hand set positions (196 vs. 1041) during this section of the video. When we include the ellipses

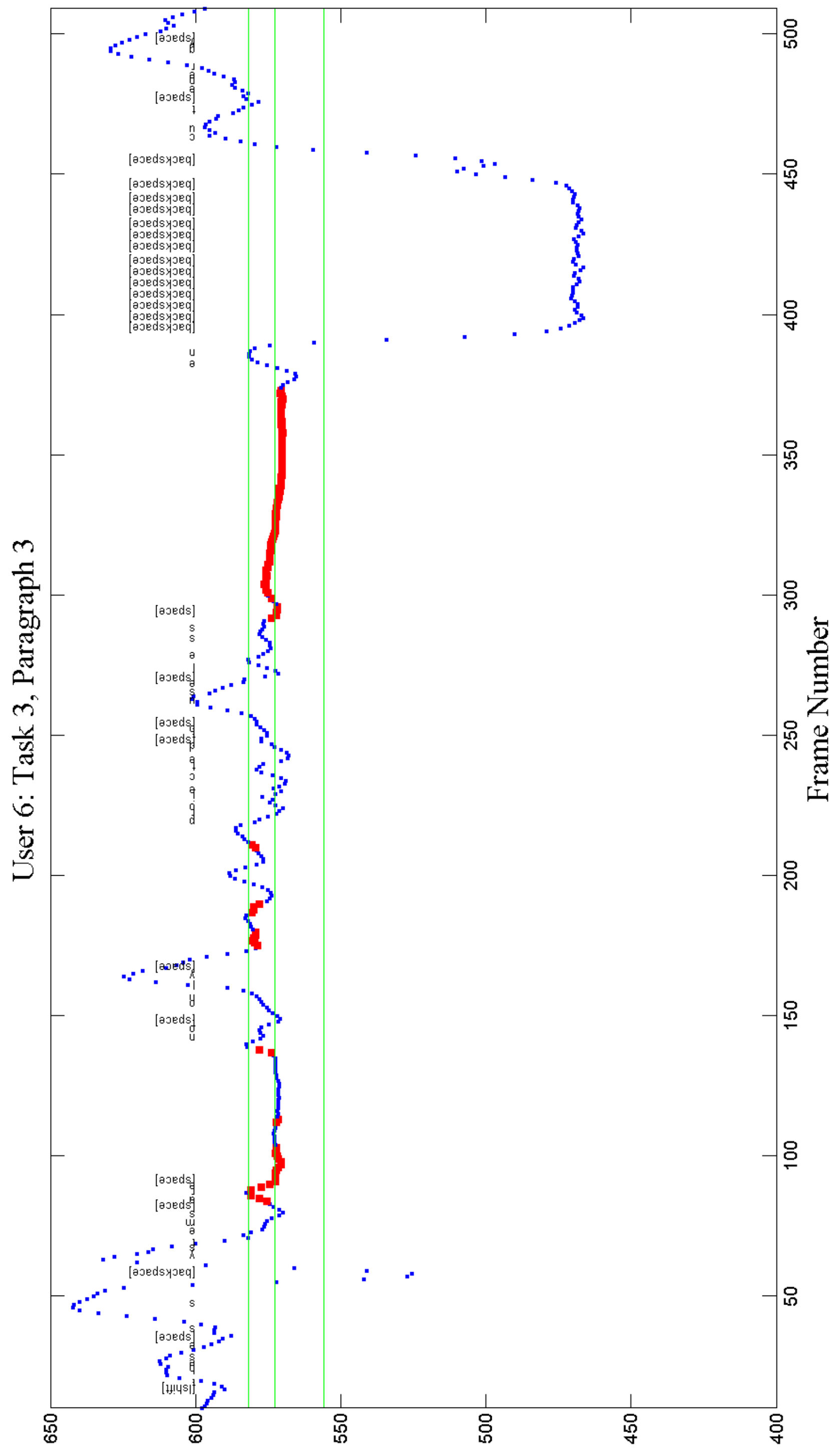


Figure 15. X coordinate of the right hand ellipse centroid for User 6 during the first approximately 500 frames of Task 3, paragraph 3. Overlaying the data are the key strokes. Combining this information shows that a large drop in the X coordinate might indicate that the user is pressing the backspace key.

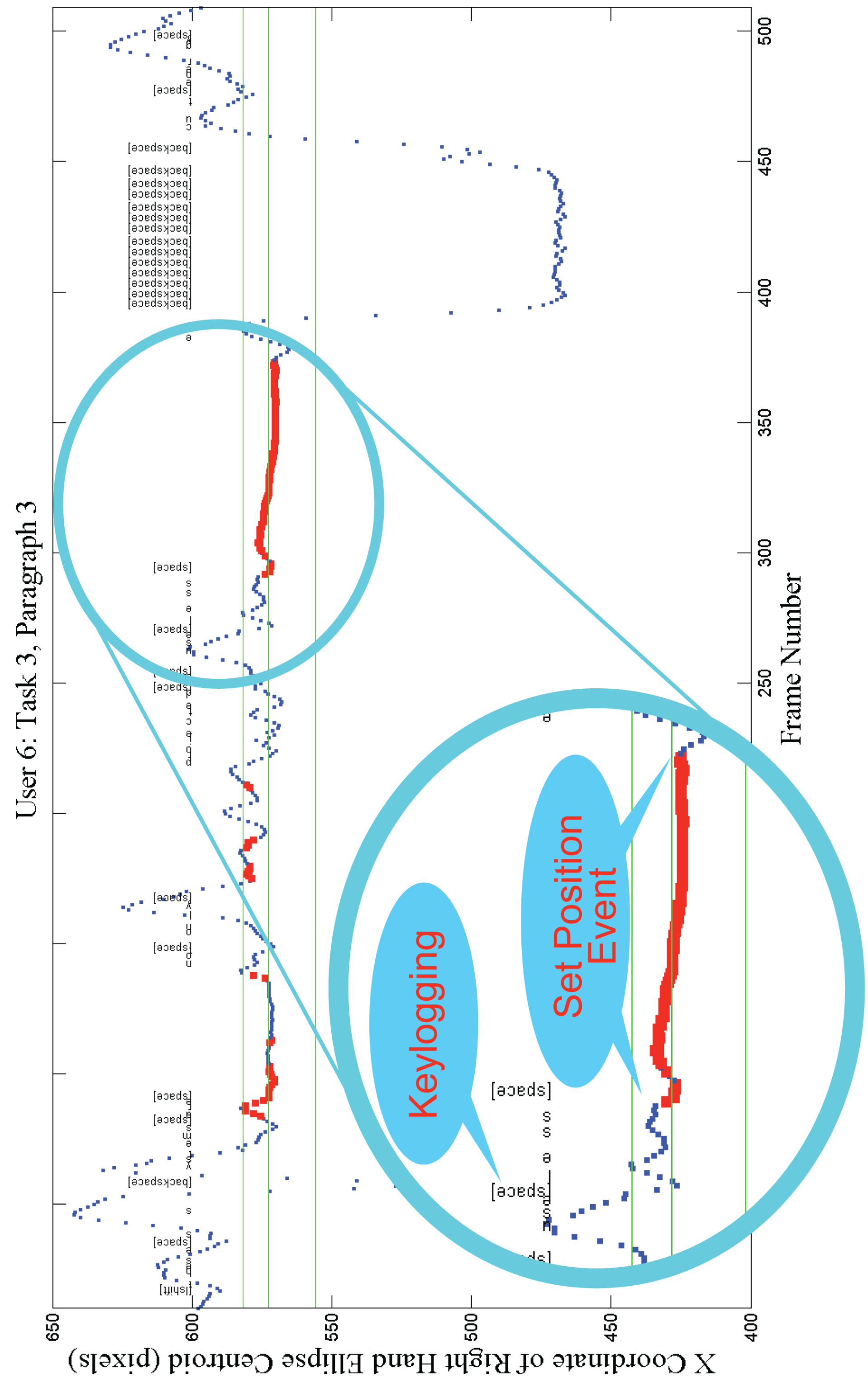


Figure 16. An enlargement of the fused keylogging and set position data.

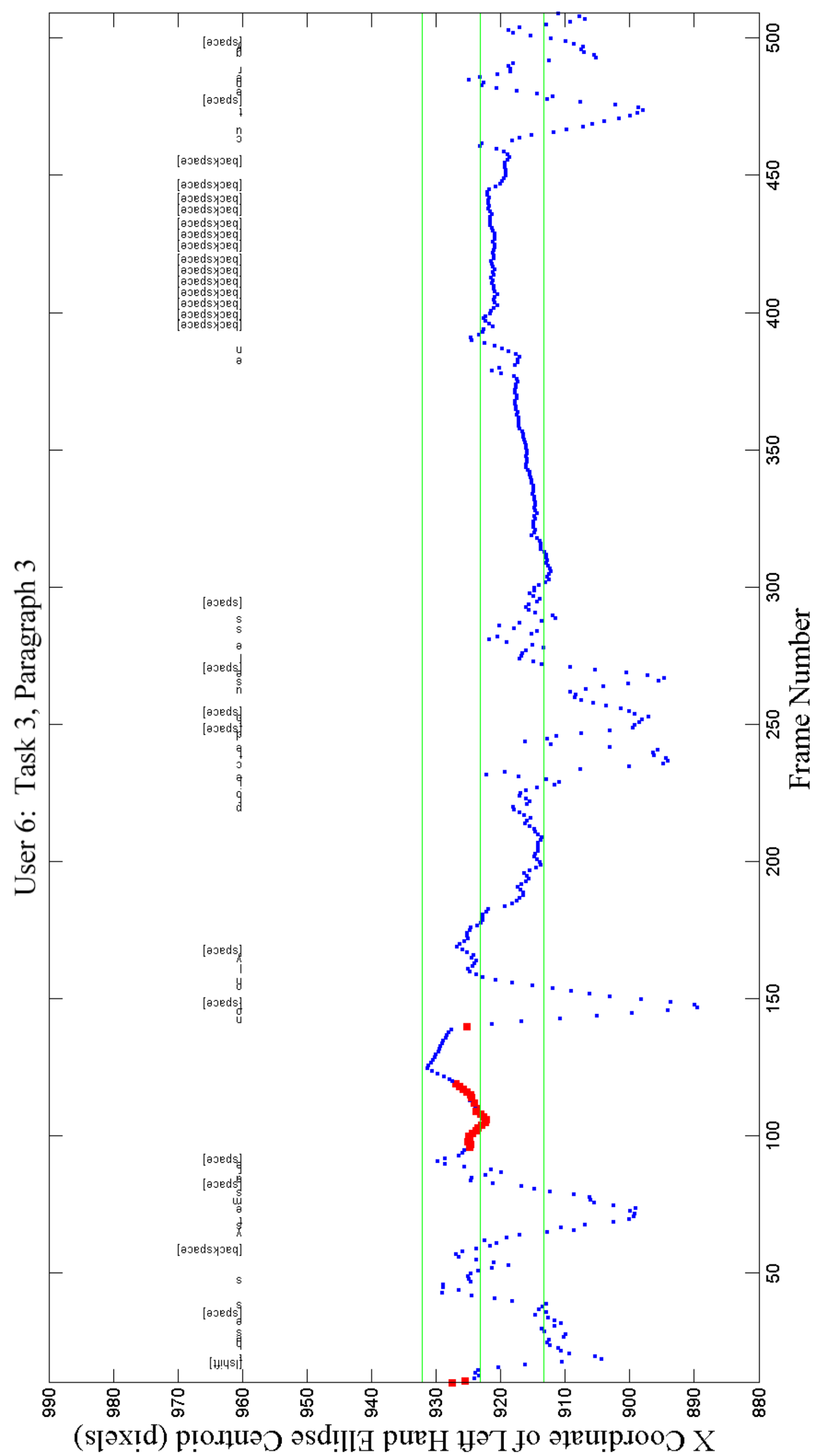


Figure 17. X coordinate of the left hand ellipse centroid for User 6 during the first approximately 500 frames of Task 3, paragraph 3. Overlaying the data are the key strokes. Frames 300-375 are not identified as set positions, when in fact they should be. Note that the X coordinate feature supports the case for a set position in that frame range.

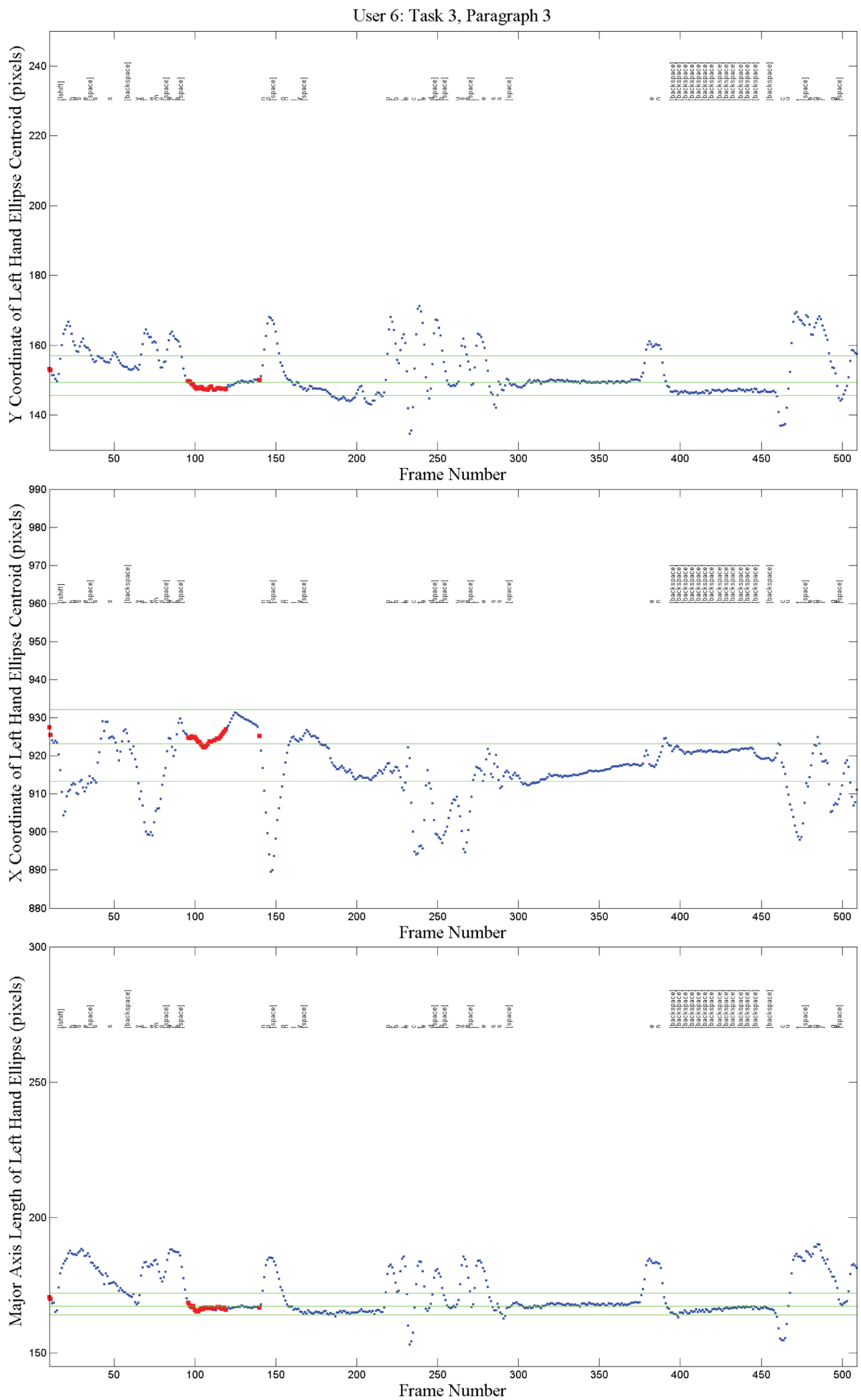


Figure 18. User 6, ellipse properties of Centroid Y-Coordinate, Centroid X-Coordinate, and Major Axis for Task 3, Paragraph 3.

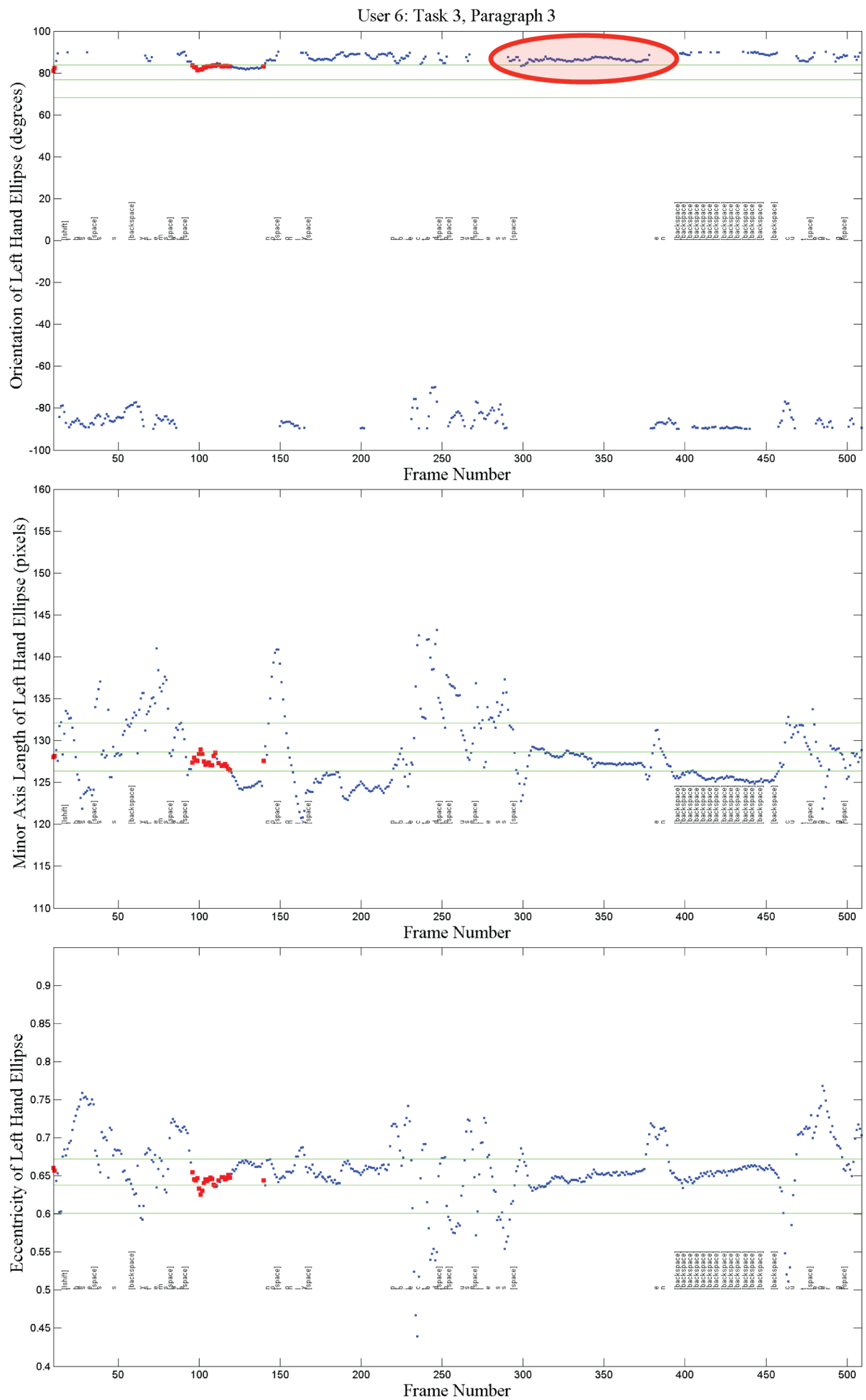


Figure 19. User 6, ellipse properties of Orientation, Minor Axis, and Eccentricity for Task 3, Paragraph 3. Noted are the values for Orientation, which are above the database range.

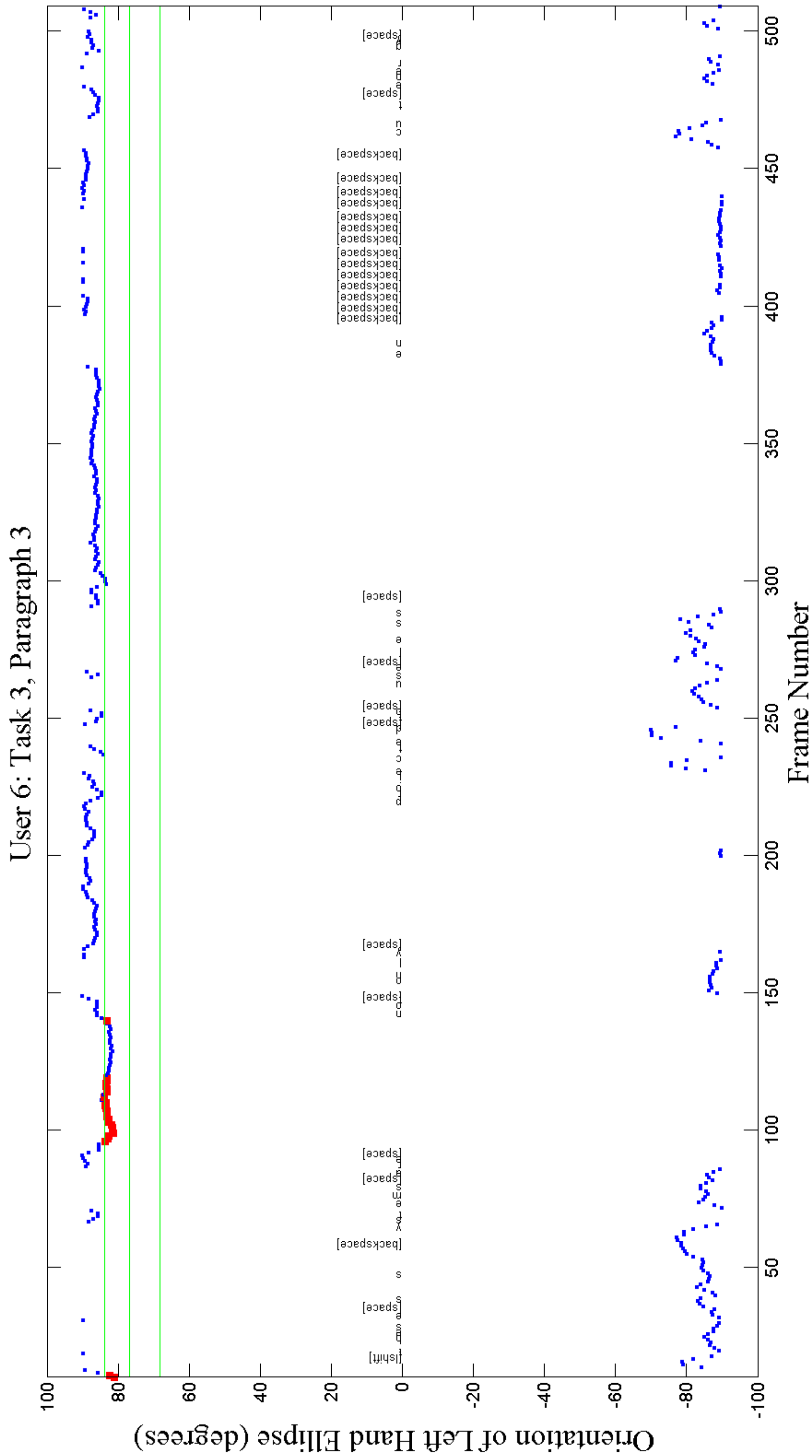


Figure 20. Orientation of the left hand ellipse for User 6 during the first approximately 500 frames of Task 3, paragraph 3. The left hand is in the set position during frames 300–375. However, the video frames used to define the set position for User 6's left hand defined the maximum positive orientation to be 83.8308 degrees, which is lower than the left hand's orientation in frames 300–375.

for just 6 frames (frames 350–355) for the left hand into the set position database, the program readily identifies many more set positions that it had previously missed (see Figure 21). Table 6 shows that for Paragraph 3, 515 set positions are now identified compared to the earlier 196 set positions for User 6's left hand. This results suggests a graceful degradation of the modeled set position in the left hand occurred in User 6's higher level work — one that may relate to posture — and not a disruptive change.

Table 6. Accuracy of Detection for User 6 During Task 3, Paragraph 3 after Set Positions with Higher Orientations are Added to Left Hand Database

Current Typist: User 6, Task 3, Paragraph 3

Labeled User	Number of Detections	Number of Confused Detections	Number of Unique Detections
Left Hand: 6	515		
8	1	0	1
9	0		
10	0		
13	0		
14	0		
15	0		
16	0		
17	0		
23	0		
Total Detections:	516		
6	1041		
8	0		
9	12	0	12
10	5	0	5
13	0		
14	0		
15	0		
16	0		
17	0		
23	0		
Total Detections:	1058		

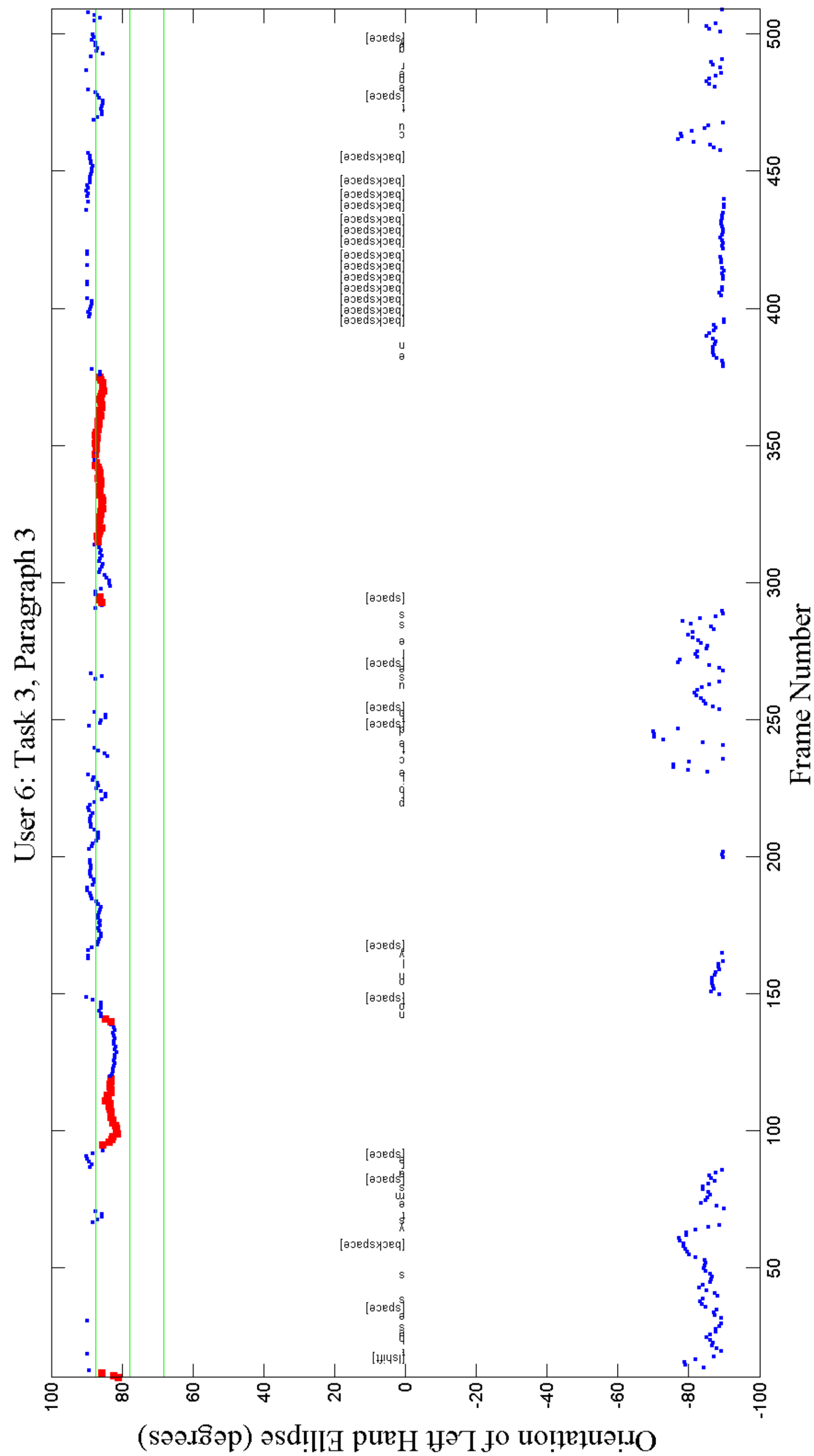


Figure 21. Orientation of the left hand ellipse for User 6 during the first approximately 500 frames of Task 3, paragraph 3, after the frames 350-355 were added to the left hand's set position database. With just those 6 frames added, almost the entire span of time that the left hand is in its set position during frames 300-375 is detected. The new maximum positive orientation for the left hand is 87.3544 degrees.

6.6 Graph Features

Because the left and the right hands can enter into their respective set positions independently, the left and right hands should be treated as separate modalities. These two modalities hold great potential due to the ease at which they can be generated using synchronized data. Most behavior biometrics take minutes to model a user and several more minutes to collect sufficient data trends for verification purposes. In contrast, the set position models for each hand can be generated from keylogging pauses and corresponding video events using on average 100-300 frames, less than 3-10 seconds of data. We have shown that a model of appropriate sensitivity can operate effectively over the course of a complex, free form task. By tracking events separately between the two hands, we were able to identify issues in the model based on apparent imbalances of detection based on less than a minute of data. Our ability to investigate the cause of the discrepancies — note, orientation, not size — may help us separate circumstances of user exhaustion (which affects posture) from user compromise.

We expect to see some discrepancies in the hands overall and as the user's workload increases. Subtleties between the hands include the assignment of responsibilities such as the manipulation of certain keys (control, shift, space, return, delete) on the keyboard and their influence on right and left hand orientation. In Figures 20 and 21, the transitions between negative and positive hand orientation can be seen in the data jumps. Although these discontinuities are not the standard way to portray angle orientation of the hands, we prefer this visualization because it distinctly shows an interesting event — the moment when a person's hand changes posture from inward oriented (toward the center of the keyboard) to outward oriented (to the edges of the keyboard), essentially from a more natural pronated posture to a less natural supinated posture. The inherent definition of the *regionprops* Orientation property is

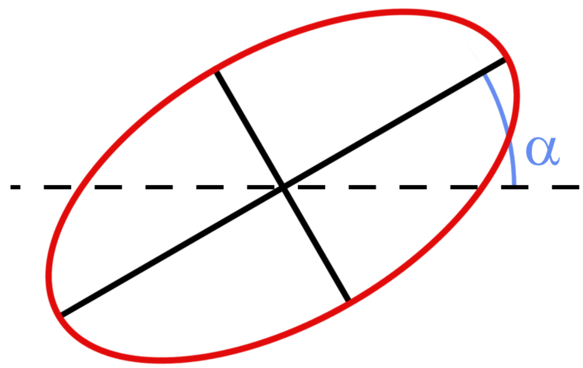


Figure 22. Orientation as defined in Matlab's *regionprops*, where α is the smallest angle between the horizontal and the Major Axis of the ellipse.

the smallest angle created between the horizontal and the Major Axis (See Figure 22). Once the Orientation feature passes through the vertical (90 degrees or $\frac{\pi}{2}$ radians), there is a sign change from positive to negative, highlighting the change in posture in the hand. 90 degrees is oriented along the vertical axis. The positive region is from 0 to $\frac{\pi}{2}$ and the negative region is from $\frac{\pi}{2}$ to π.

We can identify the keys being typed by observing the changes in the ellipse properties. We can clearly see in Figure 15 that a large decrease in the x-coordinate of the right hand indicates that 'BACKSPACE' is being pressed. For consistent orientation, the coordinate system used for the keyboard has its origin at the bottom right of the keyboard, when viewed as a typist, and increases up and to the left. Dependent on the keyboard layout, frequently used keys like BACKSPACE and RETURN may also have an associated pose. Because both of the poses for the BACKSPACE key and RETURN key tend to be in the supinate posture, we'd expect their associated poses to be more transient and rare than the set position.

6.7 Application of Results and Their Limitations

Since we have seen confirmed instances of missed set positions that were actually occurring during User 6's work, we know we likely missed other set position events elsewhere. This research was an attempt to define a simple model to identify a commonly occurring posture and to use that posture to distinguish between users. The results show this model to be workable and potentially important given the ease at identifying events of interest using synchronized data, the small amounts of data needed to generate a reasonably robust model for hand set position, the simplicity of the model, and the ability to diagnose deviations from user-centric expectations. Given their regularity, we can miss some set events and still provide verification reliably. To ensure robustness and completeness, more thorough analyses are needed to capture the full range of set positions that a computer user will enter into over the course of typing a document.

This research made no attempt to distinguish between possible changes in a computer user's set position due to fatigue or other factors, and we expect that over a typing session, as a user experiences fatigue, lack of interest, or other emotions, that their right and left set positions will change. For instance, fatigue or workload may have contributed to the deviation of the orientation feature in the left hand of User 6 discussed earlier. A more thorough study involving standard measures of fatigue — for example, skin temperature and heart rate — is warranted. Establishing separate set position models for a computer user under different operating conditions may prove more accurate in distinguishing that user and, even more desirable, in distinguishing a user's state of mind — rather than using a single set position model to distinguish a user under all conditions.

Since the right and left hands may be treated as separate modalities, the ratio of left to right hand set positions may be a distinguishing factor. Additionally, the

ratio of left to right ellipse properties, or combinations of those properties, may add robustness to this method.

Also of interest are ratios comparing QWERTY and DVORAK keyboard users. Since QWERTY users type only 32% of their strokes on the home row on a keyboard, and DVORAK users type about 70% of their strokes on the home row, we expect that more set positions would be seen in a DVORAK typer than a QWERTY typer. We found — regardless of keyboard variant — typing on the home row results in hand postures very close to the set position of a traditional typist. QWERTY users also type more strokes with the left hand, where DVORAK users type more strokes with the right hand, therefore, we might expect to see a difference in the ratio of set positions seen between left and right hands when comparing QWERTY and DVORAK users.

6.8 Verb Style Metrics as an Additional Modality

As the pool of users grows, we will have a more difficulty distinguishing between people because we will discover people who share similar anatomy, and therefore, similar set positions. In fact, we know that several different modalities are required to continue authenticating a user because each modality will naturally have a range in which it is useful. In addition, each application-relevant modality will add another layer of authentication certainty, making a user increasingly difficult for an impostor to imitate.

The set position is only one way to differentiate between users. In our initial study, there were few set positions that were mistaken for incorrect users (42 out of 1618 correct left hand set positions, and 43 out of 1659 correct right hand set positions). While set positions appear to be a good way to differentiate between ten different users, a larger pool of users will generate confusion where one set position

may be labeled as several different users. Once a user had moved out of a set position into typing, the ellipse describing the hand at that point in time may fall within the range of someone else's set positions, thereby being labeled as a set position for this other user. The multiple sources of confusion illustrate the limitations of using a set position to differentiate between people. Then of course once a user leaves the set position, how then do we continue to differentiate between people?

One possible method to continue differentiation may be to analyze the verb content and writing style of a user. The choice of verbs a person makes and the way in which they phrase their writing may be unique enough to help differentiate between people along with the set position. The modalities support each other because analysis verb style is employed when the user has moved out of the set position and is actively typing.

Since we expect to certain modalities to give a view into a person's expertise, we first take a distant view of 9 of the 10 users' documents, and can quickly observe apparent expertise with a task. Figure 23 shows a mosaic of the documents produced for Task 3. Mosaics for Tasks 1 and 2 are included in the Appendix. In Figures 26, 27, and 23, the person in each block remains the same between mosaics. These mosaics show an objective view of expertise, where the structure of a document conveys a bit of its complexity without reading the actual words. The subjects brought a range of expertise into the study. Not everyone knew how to perform a cost/benefit analysis, and those who did had varying opinions on how to construct one. People who didn't know how to make a cost/benefit analysis tended to have the bland reports without apparent structure from this distance — which can be seen in Figure 23. Their documents, quite simply, have less variation in paragraph structure. Experienced people present findings with more structure, and that structure varied among the experienced subjects.

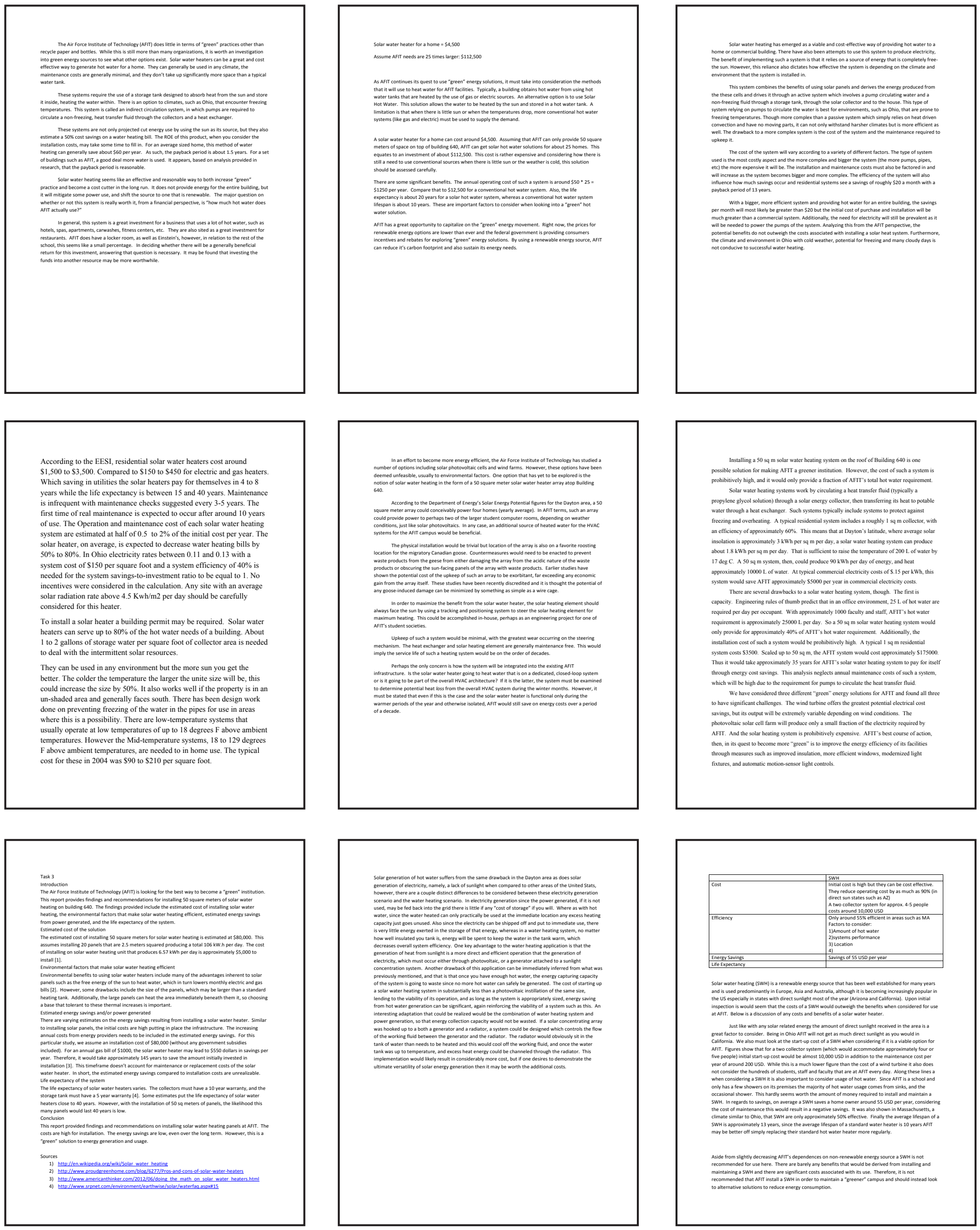


Figure 23. Mosaic of 9 of 10 subjects' documents for Task 3. Cost/benefit documents that experienced people produce tend to have more structure, and documents of inexperienced people tend to have more generic-looking paragraphs.

The way a document is prepared is a reflection of a person's expertise. People with experience may be drawn to documents with more structure and, we expect, to the structures with which they have the most familiarity.

What we can learn from this type of distant view of these documents is that people without expertise don't show identifying preferences because they simply don't have them yet. People who were experts in cost/benefit analysis showed their preferences in the way they organized their reports. We also see that, as the study wore on, some people kept their structure, and some resorted to more bland structure. This change may be due to possible loss of interest or to fatigue.

Figure 23 shows a distant view of the document's structure, but to further understand a person's expertise and combine the modalities examined in our study, we must present a more thorough analysis of the writing, and we continue by moving down from paragraph structure to the verb clause level.

Our initial concept of how to combine the video, keylogging, and verb style modalities is shown in Figure 24. Here, we have taken the highest level paragraph from User 6 — Task 3, Paragraph 3 — and constructed a verb and set position tree. Figure 25 shows the User 6, Task 3, but for Paragraph 5, which was deemed the lowest level paragraph in Task 3 for User 6. In these diagrams, the vertical axis identify the main verb per sentence in boxes. The horizontal axes show the additional verbs in the sentence in boxes. Pauses are indicated in the circles: for example, 'Pause-R?,L' indicates a long set position detected in the left hand and a likely pause in the right, where no set position was detected; 'Off-screen' indicates pauses where the user removed their hands from the keyboard. These off-screen pauses could be instances when the user is doing offscreen work — using the mouse, researching online, or switching applications — but the exact activity cannot be determined from the specific data analyzed here. Times are indicated in frames and seconds.

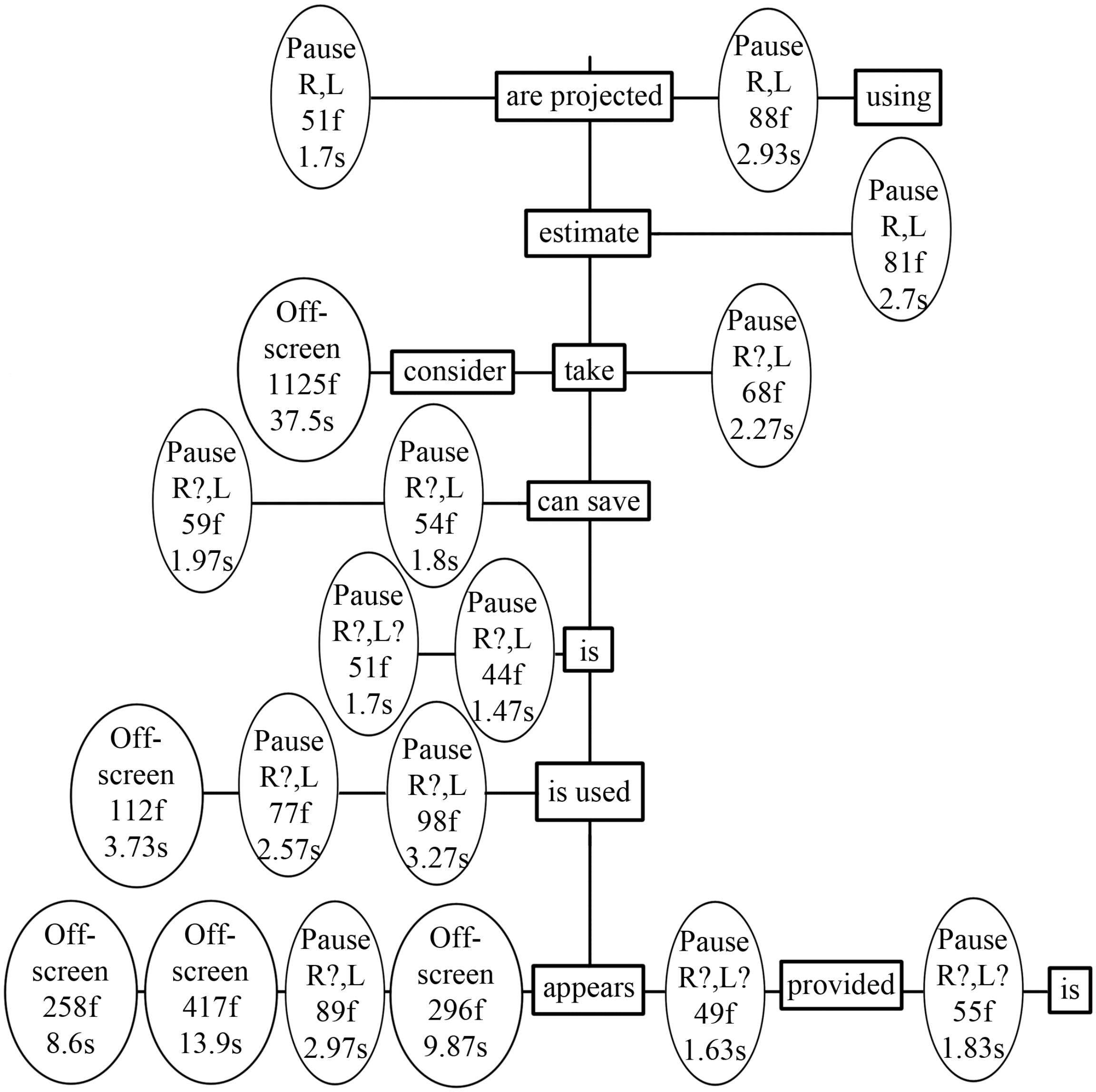


Figure 24. Verb clauses and set position for User 6: Task 3, Paragraph 3. The vertical axis shows the main verb per sentence in boxes. The horizontal axes show the additional verbs in the sentence in boxes. Pauses are indicated in the circles: for example, Pause-R?,L indicates a long set position detected in the left hand and a likely pause in the right, where no set position was detected; Off-screen indicates pauses where the user removed their hands from the keyboard. Times are indicated in frames and seconds.

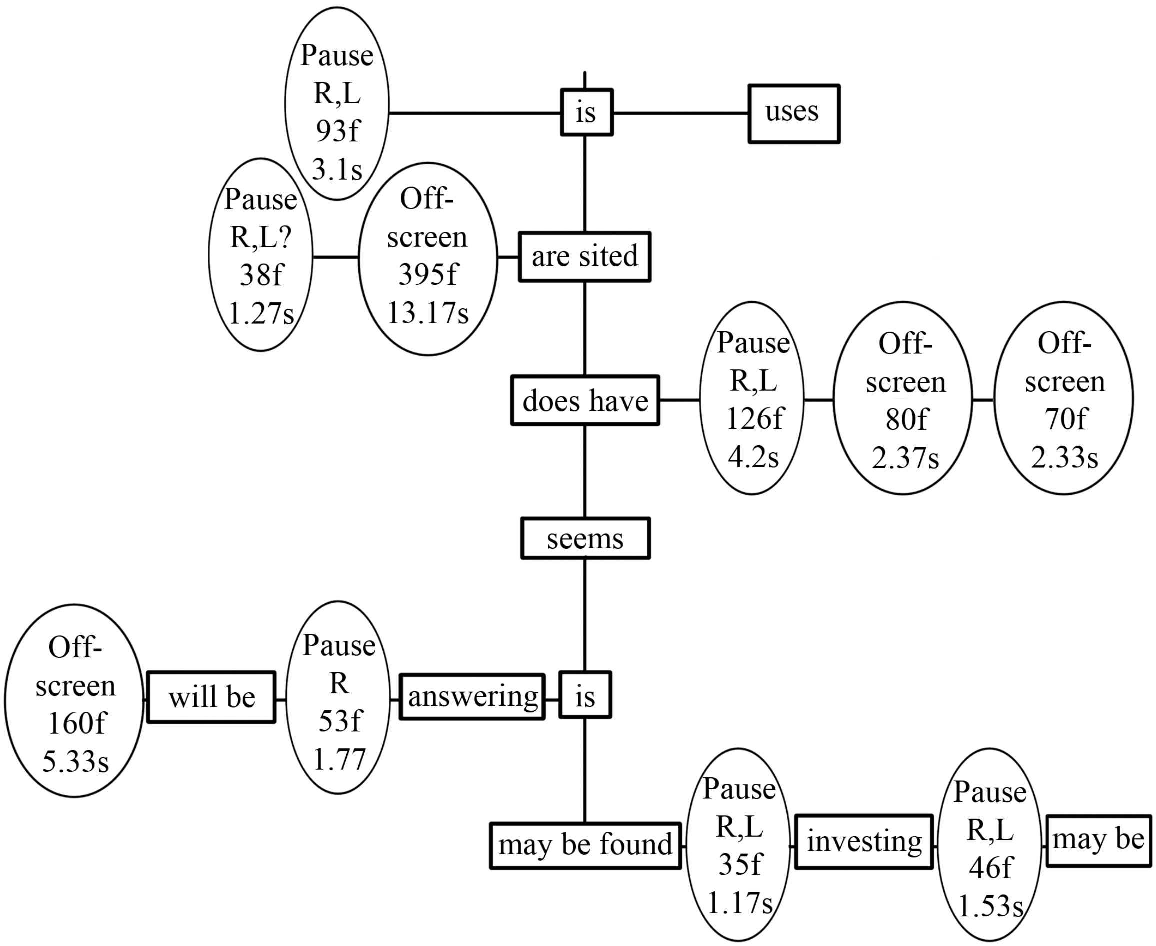


Figure 25. Verb clauses and set position for User 6: Task 3, Paragraph 5. The vertical axis shows the main verb per sentence in boxes. The horizontal axes show the additional verbs in the sentence in boxes. Pauses are indicated in the circles: for example, Pause-R?,L indicates a long set position detected in the left hand and a likely pause in the right, where no set position was detected; Off-screen indicates pauses where the user removed their hands from the keyboard. Times are indicated in frames and seconds.

An initial comparison between Figures 24 and 25 indicates that there appear to be many more pauses during higher level work. Additionally, the average duration of pauses appears to be longer during higher level work — 5.69s for paragraph 3 compared to 3.62s for paragraph 5. Also of note was the duration of time: Paragraph 3 was typed in 8023 frames, or about 4 minutes, 27.42 seconds. Paragraph 5 was typed in 4525 frames or 2 minutes 30.83 seconds. The length of time taken for each paragraph indicates that more thought was put into Paragraph 3, including possible calculations, inferred from the verb 'estimate'. This time period was where User 6 was doing most of the analysis, operating at Bloom's Taxonomy levels of Application and Analysis, as indicated by the verbs 'consider', 'appears', and also 'estimate'. In contrast, Paragraph 5 has more verbs indicating Comprehension or Recall may be occurring — 'are sited' and 'does have'.

6.9 Summary

Although these results cover just a small subset of the data gathered, they are promising and warrant further investigation into just how users may differentiate themselves while performing their highest level work. Most importantly they demonstrate a positive outcome: We can expect more frequent pauses in higher level work. The modalities tracking right and left set positions are thus more likely to be effective in high level work as long as the subject's set position posture does not alter significantly. The instances when a user is performing Application and Analysis offer a view into the user's particular preferences and thus identifying behavioral characteristics.

VII. Conclusions

This thesis examined the use of elliptical features to model of a user's set position in order to differentiate between computer users. We investigated the fusion of data features extracted from video, with keylogging and text. We can differentiate between computer users via this neutral hand posture with only a few seconds of training data, using an overhead camera for sensing. This sensitive and specific measurement is consistent throughout typing in a free form task involving internet searches and a cost benefit analysis. By fusing this video data with a Bloom's Taxonomy analysis of typed text and keylogging data, we have developed a method to determine the level of work performed and showed that a computer user may be differentiated by this neutral hand posture even during complex work — where they are most likely to reveal preferences. The set positions of each hand and the user's apparent competency all serve as individual modalities that can serve in the act of authentication.

Activities indicating more thought and Application/Analysis level of work can point to the expertise of the user, and is where we are most interested. We theorize this type of activity is where people distinguish themselves most, and therefore, it is the most important activity to recognize. Additionally, during typing, work at this higher level appears to have more instances of 'set position' than lower level work and also offers additional means to verify a user once they leave the 'set position'. The way a user behaves during higher level work or under stress needs to be thoroughly examined and mined for distinguishing modalities so that a computer system can continue to authenticate the user at their most productive state.

The findings of this research contribute directly to biometrics. We have created a model that functions while a person is in direct interaction with an object. The 'set position' can be applied to next generation touch screen devices. These smart devices will be able to take advantage of our advanced understanding of psychomotor

behavior and customize interfaces to make the device easier to use for the primary user but harder for others. Although the layout of these touch screen devices differ from that of a keyboard, a similar 'set position' may be found that is comparatively unique to each user, allowing one method of authentication.

Connecting behavior with proficiency will enable us to refine our assessment of human authority — that mix of competency and influence needed to get good work done. This connection gives us a method of identifying experts, novices, and certain threats by their subtle interactions with the environment.

7.1 Future Work

Future work will build upon the simple model developed here, adding fingertip tracking. We will use inverse kinematics via the Groebner Basis Theory approach [26] [27] to create an accurate hand model that more precisely captures the posture of the hands. Passive radar imaging may give us the ability to see hands grasping an object without fear of occlusion.

In future studies, we will apply a more automated method of extracting the hand from the background. Currently employed was a hard coded RGB value range, within which a given pixel was determined to be skin. This type of coding is insufficient, as skin color changes based on lighting conditions and ethnicity. YCbCr color space, which separates luminance from color information and is additionally independent of racial skin color [7] [9] will be investigated.

We will continue data fusion of video, text, and keylogging data to model how a user behaves when doing their most compelling work. We want to characterize competency and recognize when a user is performing at a higher level of competence. Future environments for study may include a more variable, competitive setting such as the ACE Hackfest [32], an annual large-scale cyber warfare exercise held at AFIT.

Appendix A. Error Matrices

Below are tables containing the confused and unique detections for ten users for the left and right hands.

Table 7. Comparison of Confused Detections Among Users for Left Hand

	# Confused Detections During Respective Typists									
Left Hand	6	8	9	10	13	14	15	16	17	23
Current Typist: 6	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	2	0	0	0	0	0	0
10	0	15	25	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0

Table 8. Comparison of Confused Detections Among Users for Right Hand

	# Confused Detections During Respective Typists									
Right Hand	6	8	9	10	13	14	15	16	17	23
Current Typist: 6	0	0	0	0	0	0	0	0	0	0
8	0	0	31	0	0	0	0	0	0	0
9	0	1	0	11	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0

Table 9. Comparison of Unique Detections Among Users for Left Hand

	# Unique Detections During Respective Typists									
Left Hand	6	8	9	10	13	14	15	16	17	23
Current Typist: 6	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	15	0	0	0	0
9	0	9	0	8	0	0	0	0	0	0
10	0	18	6	0	0	0	0	0	4	0
13	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	2	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
17	0	48	0	3	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0

Table 10. Comparison of Unique Detections Among Users for Right Hand

	# Unique Detections During Respective Typists									
Right Hand	6	8	9	10	13	14	15	16	17	23
Current Typist: 6	0	0	0	0	0	0	0	0	0	0
8	0	0	55	0	0	0	0	0	0	0
9	0	0	0	5	0	0	0	0	0	0
10	0	1	7	0	0	1	0	0	1	0
13	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	7	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	1	0	0	0
17	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0

Appendix B. Active Learner Scavenger Hunt

In the human study conducted for this thesis, each subject enacted a computer based scavenger hunt. The scavenger hunt required the participant to write a short essay providing a cost-benefit analysis. Subjects were expected to have various levels of skill in typing and in the formatting and preparation of a cost-benefit analysis. We chose the topics that subjects were not expected know well so that there would be a learning aspect to the task. The three tasks are provided on the following three pages in the manner that they were presented to the subjects.

Task 1

Being “green” can involve several different facets. This could include using an energy source that is sustainable into the future as well as friendly to the environment such as solar, wind or tidal energy. Being “green” can also involve making changes to current architecture of a building or generating new ways to operate in order to consume less energy.

The Air Force Institute of Technology (AFIT) is looking for the best way to become a “green” institution. They need your help determining the return on investment for installing a Wind Turbine behind the facility.

A. The deliverable for this task is a ~1000 word report detailing your findings and recommendation on the best course of action for turning AFIT into a “green” campus.
    a. You should use the internet to find factual information to include in your report. Documentation of your sources does not need to occur but please do not copy and paste information directly from a web page.
    b. Factors to take into consideration when making your recommendation
        i. Estimated cost of the solution
        ii. Environmental factors that make a wind turbine efficient
        iii. Estimated energy savings and/or power generated
        iv. Life expectancy of the system

The costs and benefits may be best expressed in a table. Also, please include any other information you deem to be necessary.

After completing the report, copy it to the given removable hard drive.

Task 2

Being “green” can involve several different facets. This could include using an energy source that is sustainable into the future as well as friendly to the environment such as solar, wind or tidal energy. Being “green” can also involve making changes to current architecture of a building or generating new ways to operate in order to consume less energy.

The Air Force Institute of Technology (AFIT) is looking for the best way to become a “green” institution. They need your help determining the return on investment for installing 50 square meters of solar energy photovoltaic cells on the top of building 642.

A. The deliverable for this task is a ~1000 word report detailing your findings and recommendation on the best course of action for turning AFIT into a “green” campus.
    a. You should use the internet to find factual information to include in your report. Documentation of your sources does not need to occur but please do not copy and paste information directly from a web page.
    b. Factors to take into consideration when making your recommendation
        i. Estimated cost of the solution
        ii. Environmental factors that may make solar cells more efficient
        iii. Estimated energy savings and/or power generated
        iv. Life expectancy of the system

The costs and benefits may be best expressed in a table. Also, please include any other information you deem to be necessary.

After completing the report, copy it to the given removable hard drive.

Task 3

Being “green” can involve several different facets. This could include using an energy source that is sustainable into the future as well as friendly to the environment such as solar, wind or tidal energy. Being “green” can also involve making changes to current architecture of a building or generating new ways to operate in order to consume less energy.

The Air Force Institute of Technology (AFIT) is looking for the best way to become a “green” institution. They need your help determining the return on investment for installing for installing 50 square meters of solar water heating on building 640.

A. The deliverable for this task is a ~1000 word report detailing your findings and recommendation on the best course of action for turning AFIT into a “green” campus.
    a. You should use the internet to find factual information to include in your report. Documentation of your sources does not need to occur but please do not copy and paste information directly from a web page.
    b. Factors to take into consideration when making your recommendation
        i. Estimated cost of the solution
        ii. Environmental factors that may make solar water heating more efficient
        iii. Estimated energy savings and/or power generated
        iv. Life expectancy of the system

The costs and benefits may be best expressed in a table. Also, please include any other information you deem to be necessary.

After completing the report, copy it to the given removable hard drive.

Appendix C. Scavenger Hunt Mosaics for Tasks 1 and 2

Mosaics for Tasks 1 and 2 for the documents produced of 9 out of 10 subjects from the scavenger hunt. The person in each block remains the same between mosaics.

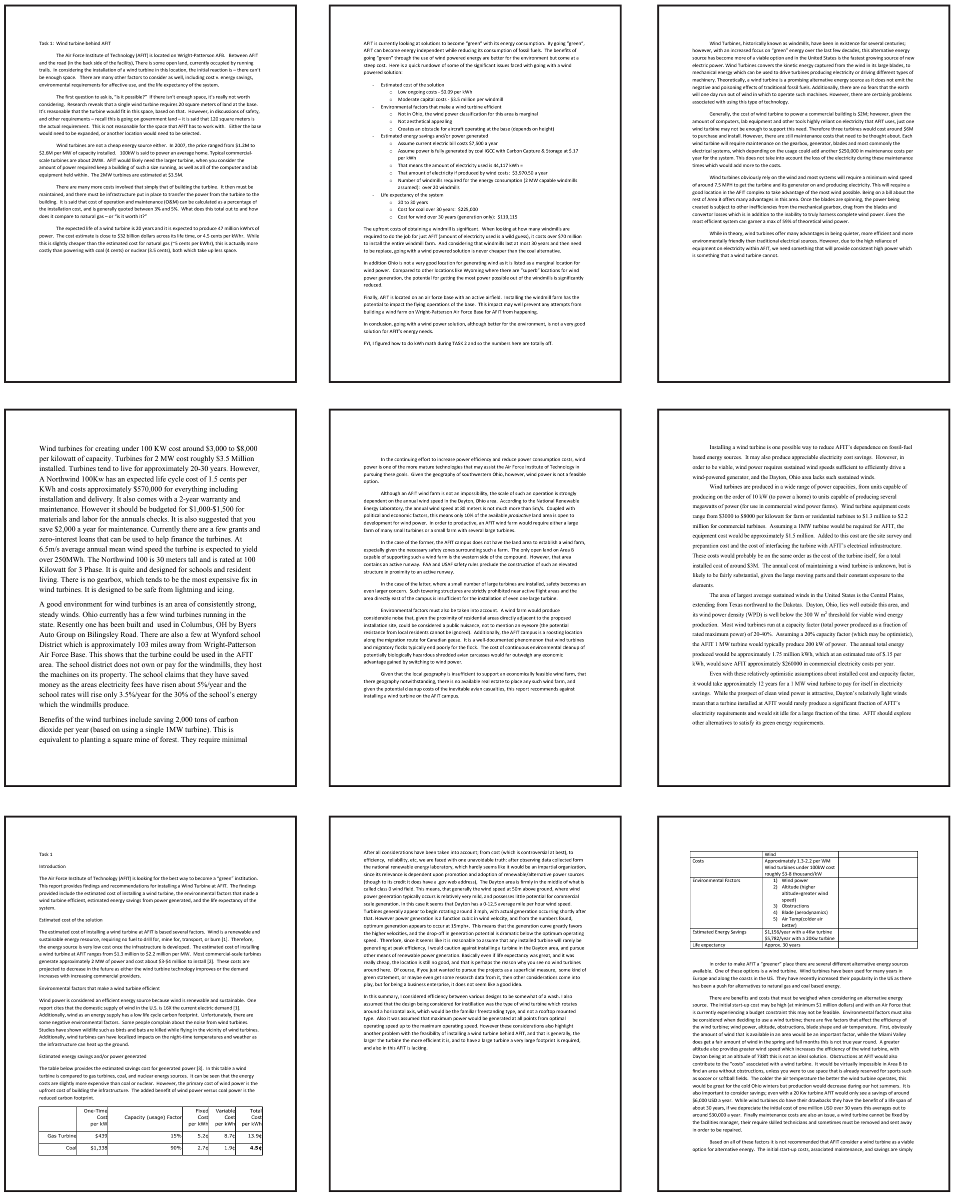


Figure 26. Mosaic of 9 of 10 subjects' documents for Task 1.

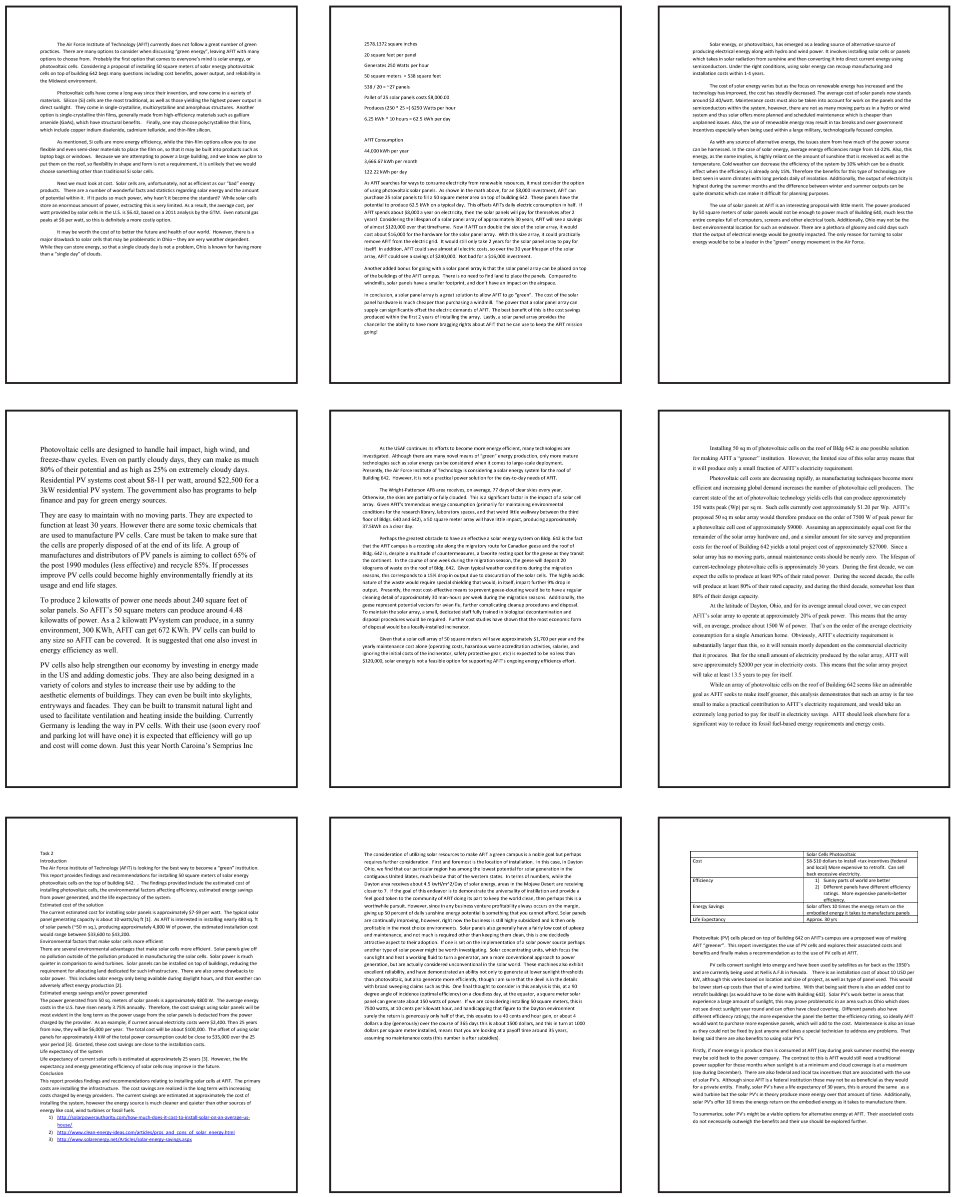


Figure 27. Mosaic of 9 of 10 subjects' documents for Task 2.

Bibliography

1. S. Wiedenbeck, J. Waters, J.C. Birget, A. Brodskiy, and N. Memon. Authentication using graphical passwords: effects of tolerance and image choice. In *Proceedings of the 2005 symposium on Usable privacy and security*, pages 1–12. ACM, 2005.

2. S. Wiedenbeck, J. Waters, J.C. Birget, A. Brodskiy, and N. Memon. Passpoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, 63(1):102–127, 2005.

3. S.K. Bandyopadhyay, D. Bhattacharyya, and P. Das. User authentication by secured graphical password implementation. In *7th Asia-Pacific Symposium on Information and Telecommunication Technologies, 2008. APSITT*, pages 7–12. IEEE, 2008.

4. Somini Sengupta. Logging in with a touch or a phrase (anything but a password). World Wide Web Page, December 2011. Keywords: logging, touch, phrase, password.

5. Andy Greenberg. Gesture-based login apps for ipad and iphone aim to banish passwords from touchscreens. World Wide Web Page, November 2011. Keywords: gesture-based, login apps, touchscreens, passwords, ipad, iphone.

6. Y. Sutcu, Q. Li, and N. Memon. Secure biometric templates from fingerprint-face features. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07*, pages 1–6. IEEE, 2007.

7. A. Ben Jmaa, W. Mahdi, Y. Ben Jemaa, and A. Ben Hamadou. A new approach for digit recognition based on hand gesture analysis. *(IJCSIS) International Journal of Computer Science and Information Security*, 2(1), 2009.

8. C. Manresa, J. Varona, R. Mas, and F. Perales. Hand tracking and gesture recognition for human-computer interaction. *Electronic letters on computer vision and image analysis*, 5(3):96–104, 2005.

9. A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. Hand tracking and affine shape-appearance handshape subunits in continuous sign language recognition. In *Proceedings of International Conference ECCV Wkshp: SGA, Heraklion, Crete*, volume 1, 2010.

10. KA Barhate, KS Patwardhan, S.D. Roy, S. Chaudhuri, and S. Chaudhury. Robust shape based two hand tracker. In *International Conference on Image Processing, 2004. ICIP'04. 2004*, volume 2, pages 1017–1020. IEEE, 2004.

11. T. Rhee, U. Neumann, and JP Lewis. Human hand modeling from surface anatomy. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 27–34. ACM, 2006.

12. Wei Shu and David Zhang. Automated personal identification by palmprint. *Optical Engineering*, 37(8):2359–2362, 1998.

13. Ross A. Jain, A. K. and S. Pankanti. A prototype hand geometry-based verification system. In *In Proceedings of 2nd International Conference on Audio and Video-based Biometric Person Authentication(AVBPA)*, page 166 171, 1999.

14. S. Kim, Y. Park, K. Lim, H. Lee, S. Kim, and S. Lee. Fingertips detection and tracking based on active shape models and an ellipse. In *TENCON 2009-2009 IEEE Region 10 Conference*, pages 1–6. IEEE, 2009.

15. Stefan Stegmueller. Hand and finger tracking with kinect depth data. World Wide Web Page, 2011. Keywords: Candescent NUI, Kinect, hand, finger, tracking.

16. J. Allen, N. Chambers, G. Ferguson, L. Galescu, H. Jung, M. Swift, and W. Taysom. Plow: A collaborative task learning agent. In *Proceedings of the National Conference on Artificial Intelligence*, volume 22, page 1514. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2007.

17. G. Ferguson, J.F. Allen, et al. Trips: An integrated intelligent problem-solving assistant. In *Proceedings of the National Conference on Artificial Intelligence*, pages 567–573. JOHN WILEY & SONS LTD, 1998.

18. M. Swift, G. Ferguson, L. Galescu, Y. Chu, C. Harman, H. Jung, I. Perera, Y.C. Song, J. Allen, and H. Kautz. A multimodal corpus for integrated language and action. In *Proceedings of the International Workshop on MultiModal Corpora for Machine Learning*, 2012.

19. Video and image retrieval and analysis tool (virat), 2008. DARPA-BAA-08-20.

20. Defense Industry Daily. Darpa's virat: Video search, with a twist. World Wide Web Page, September 2010.

21. Active authentication, 2012. DARPA-BAA-12-06.

22. Kyle Bailey. Computer based behavioral biometric authentication via multi-modal fusion. Master's thesis, Air Force Institute of Technology, March 2013.

23. Geoffrey Holmes Bernhard Pfahringer Peter Reutemann Ian H. Witten Mark Hall, Eibe Frank. The weka data mining software: an update. *SIGKDD Explorations*, 11(1):10–18, 2009.

24. Ronald F. Tuttle David J. Bunker Christoph C. Borel Kimberly Kendricks, Anum Barki. An inverse-kinematic appraoch using groebner basis theory applied to gait cycle analysis. 2012.

25. Anum Barki. An inverse kinematic approach using groebner basis theory applied to gait cycle analysis. Master's thesis, Air Force Institute of Technology, March 2013.

26. Kimberly Kendricks. *Solving the Inverse Kinematic Robotics Problem: A Comparison Study of the Denavit-Hartenberg Matrix and Groebner Basis Theory*. PhD thesis, Auburn University, August 2007.

27. Kimberly D. Kendricks, Adam M. Fullenkamp, Robert Mcgrellis, Jonathan Juhl, and Ronald F. Tuttle. An inverse kinematic mathematical model using groebner basis theory for arm swing movement in the gait cycle, July 2010.

28. Amy L. Magnus and Mark E. Oxley. Arrogance in classification. In *Proceedings of the 2003 IEEE Aerospace Conference*, volume 5, pages 2037 – 2048, Big Sky, Montana, March 2003.

29. Amy L. Magnus and Mark E. Oxley. Information forensics and the art of inquiry. In *Proceedings of SPIE, Intelligent Computing: Theory and Applications IV*, volume 6229, Orlando, FL, April 2006.

30. D. R. Clark. Bloom's taxonomy of learning domains. World Wide Web Page, June 1999. Keywords: Bloom, taxonomy, learning domain.

31. Elizabeth. J. Simpson. *The Classification of Educational Objectives in the Psychomotor Domain*. Gryphon House, 1972.

32. Jacob Ames. Advanced course in engineering cyber security boot camp 2010. World Wide Web Page, April 2010. Keywords: ACE Hackfest, Advanced Course in Engineering, Cyber Security Boot Camp.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*
21-03-2013

2. REPORT TYPE
Master's Thesis

3. DATES COVERED *(From — To)*
Sept 2011 — Mar 2013

4. TITLE AND SUBTITLE
DISCRIMINATION OF NEUTRAL POSTURES IN COMPUTER BASED WORK

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Keith, Alanna, Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENP-13-M-19

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 N. Randolph St, Suite 325, Room 3112
Arlington VA 22203
DSN 426-7796, COMM 703-696-7796
Email: tristan.nguyen@afosr.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL,RF

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT A biometric system recognizes users based on the way they physically interact with the system. In this work, we discover a common behavior that a typist consistently displays in non-trivial computer work. We sought to demonstrate three objectives: first, compelling proof that a user can be actively recognized over the course of a lengthy task via a neutral posture struck multiple times in that task; two, a sensing concept for capturing the neutral posture, and, third, an objective method for determine the level of work performed by each typist.
This thesis develops a model for hand tracking using a simple ellipse to describe the neutral posture where a typist pauses before typing. Initial results of a group of 10 users indicate that the neutral posture can be established with only a few seconds of training data and can perform with approximately 92.1% accuracy. Analysis of the typed text determined the complexity of the typists' work using Bloom's Taxonomy - a taxonomy based on verb usage; parsed verb phrases indicated the level of competency that the users endeavored to demonstrate. This competency or expertise may further distinguish users and their performance in their most engaging work.

15. SUBJECT TERMS
Biometric Authentication System, Hand Tracking, Bloom's Taxonomy, Set Position, Neutral Position, Competency

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	U	96	Dr. Amy L. Magnus, AFIT/ENP
					19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4555; amy.magnus@afit.edu